University of Dundee

DOCTOR OF PHILOSOPHY

Nonlinear Feedback and Stochastic Control in Translation

Wang, Liang

Award date:
2015

Nonlinear Feedback and Stochastic Control in Translation

By

Liang Wang

A Thesis submitted for the degree of Doctor of Philosophy

Division of Mathematics

University of Dundee

Dundee

June 2015

Contents

List of Figures

List of Tables

Declaration

I declare that the following thesis is my own composition and that it has not been submitted before in application for a higher degree.

Liang Wang

Certification

This is to certify that Liang Wang has complied with all the requirements for the submission of this Doctor of Philosophy thesis to the University of Dundee.

Prof. Fordyce Davidson

Abstract

Proteins are large biological molecules that perform a vast array of functions within living organisms. Proteins are made from a process called *translation*, in which a ribosome decodes mRNA, a single-stranded copy of DNA, to produce a specific amino acid chain.

Given the essential role of proteins in maintaining life, it is of central importance to comprehend the translation process and how it is regulated. Translation process can be divided into three major stages: initiation, elongation and termination. Regulation can occur at any of these stages to control protein production. In most cases, regulation primarily targets the initiation stage. Another direct mechanism to ensure accurate protein level in the cell is to regulate the stability of mRNA.

Using mathematical modelling, in this thesis we investigate how protein production is controlled. We use a stochastic modelling approach called the Totally Asymmetric Simple Exclusion Process to model the translation process. Numerical simulation acts as a complementary tool. We first investigate how mRNA stability affects protein production from one mRNA during its lifetime. Next, we investigate auto-negative feedback on translation initiation and its substantial impact on controlling protein level in a cell. Finally, we incorporate ribosome recycling into the auto-negative feedback control. Novel results such as oscillation and bistable switching in protein level are found via this mathematical analysis. These predictions invite experimental testing.

Chapter 1

Introduction

This introductory chapter is employed to present basic biological and mathematical background required to clarify the main scope of this PhD project. We also outline the work presented in this thesis.

1.1 Biological Background

1.1.1 Translation of Messenger RNA (mRNA)

The central dogma of molecular biology states that genetic information is coded in self replicating DNA and undergoes unidirectional transfer to a single stranded copy of DNA called messenger RNA or mRNA in a process called *transcription*, mRNA serves as templates for protein synthesis in the related process of *translation* [32, 33]. This thesis focuses on the process of translation. A detailed description of this process is as follows.

In brief, translation is the process of manufacturing proteins. Proteins are biological

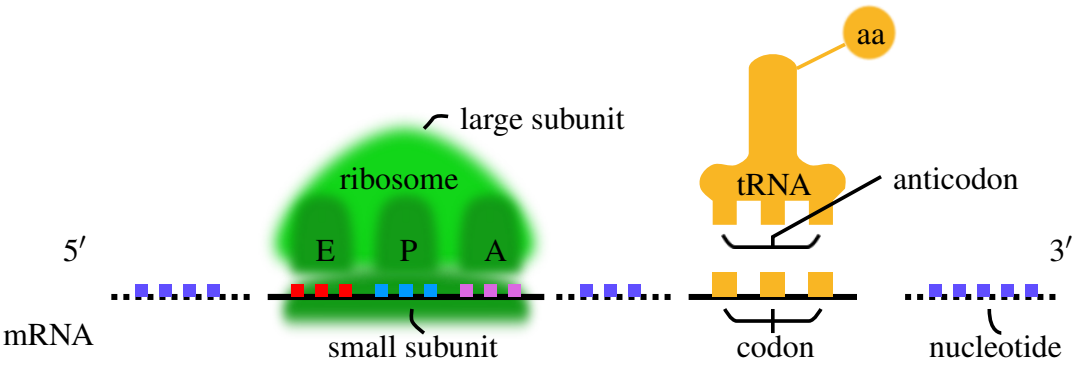


Figure 1.1: **Elemental components for mRNA translation.** Illustration of mRNA, ribosome, tRNA and amino acid.

functional molecules, comprising one or more chains of amino acids (aa) [113]. A chain of amino acids is also termed as a polypeptide. Twenty types of amino acids are used to naturally assemble proteins, and these amino acids are known as natural amino acids or proteinaceous amino acids [76, 143]. Figure 1.1 provides the illustration of the elementary components for translation. The mRNA is a chain of nucleotides with a 5' end and a 3' end. In mRNA, genetic information is encoded in a sequence of nucleotides. A, G, C, U are used to represent four different bases of nucleotide in the form of the nucleoside triphosphastes (ATP, GTP, CTP and UTP). Three nucleotides as a group is commonly referred to as a *codon*. Therefore, there are $4^3 = 64$ different codons (possible combinations of A, G, C, U) [34]. Each codon corresponds to a certain amino acid. However, most amino acids can be indicated by more than one codon. For example, codon GUC codes for amino acid Valine while Valine can also be specified by codons GUU, GUA and GUG. This phenomenon of duplication is because of the redundancy between 64 possible codons and yet only 20 natural amino acids. The first codon of an mRNA to be translated is called a start codon (in 5' region). AUG is the most common start codon [109]. Relatively, the codon that signals termination of translation is called a stop codon. There are several stop codons in standard mRNA: UAG, UAA and UGA [70, 88].

A ribosome is a complex molecule of proteins and RNA, consisting of a small subunit

and a large subunit. The ribosome serves as a machine to assemble amino acids to a polypeptides, and finally release the polypeptides as a protein [12, 120]. Ribosomes move in the 5' to 3' direction (i.e., reads the genetic information from start codon to stop codon). Ribosomes have three binding sites for transfer RNA (tRNA) designated A, P, E. Site A recruits an aminoacyl-tRNA; site P is for binding of a peptidyl-tRNA; and site E provides binding room for a free tRNA to exit. The tRNA is responsible for transporting the raw material of protein, amino acid, to ribosomes. The most prominent structure of tRNA is the so-called *anticodon*. The anticodon is utilized to recognise and decode the corresponding codon [54]. Therefore, tRNA can bind to sites in ribosome, recognise the codon and deliver the correct amino acid to the ribosome to subsequently build protein. Particularly, a tRNA bound with the correct amino acid is referred to as a 'charged' tRNA. Further information about the ribosome and tRNA operation will be provided next.

In general, both eukaryotic and prokaryotic translation processes can be divided into three main stages: initiation, elongation and termination, though the mechanism in eukaryotic cells can be more complicated [2, 13, 70, 97, 106].

The first step of the initiation stage involves assembly of a small ribosome subunit to the sequence of an mRNA called a *ribosome binding site*, or *RBS*, with the help of certain initiation factors. In eukaryotes, the ribosome binding site is the 5' cap of an mRNA. In prokaryotes, the ribosome binding site is a so-called *Shine-Dalgarno sequence* located a few nucleotides upstream of the start codon of an mRNA. Subsequently, an initiator tRNA binds to the P site and the small ribosome subunit scans the mRNA in the 5' to the 3' direction until the start codon is reached. After that, a large ribosome subunit is recruited to associate with the small ribosome subunit, leading to the formation of a complete ribosome that is ready to start translating. Subsequently, translation proceeds to the elongation stage.

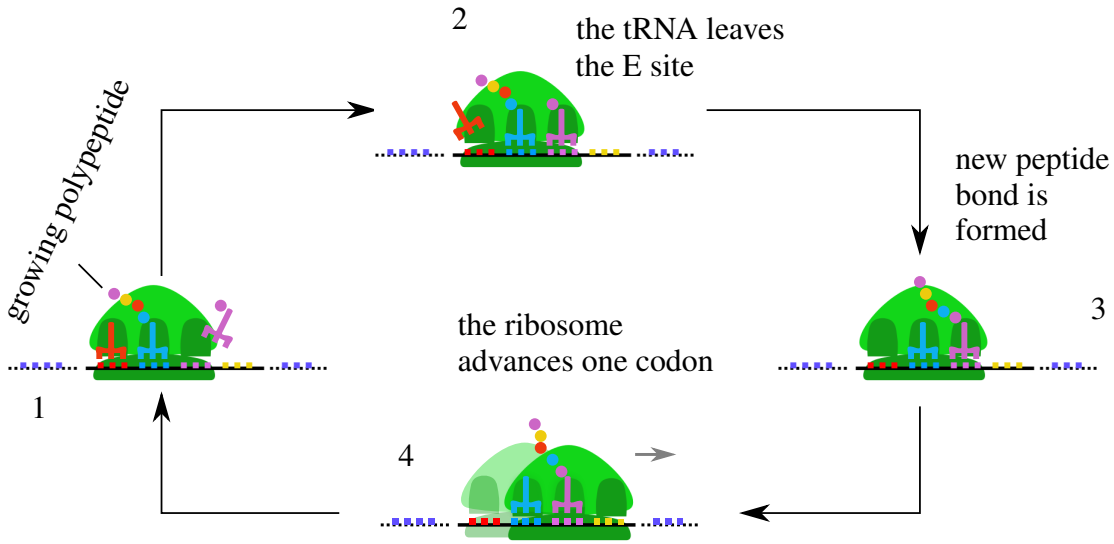


Figure 1.2: **Sketch of the elongation process in translation.** Four steps of elongation numbered consecutively by $1-4$ are displayed. Ribosome moves forward one codon by one codon. Diagram is reproduced from [27].

The sketch of the elongation process is displayed in Figure 1.2. After initiation, the initiator methionyl tRNA is positioned in the P site. The first step of the elongation is the binding of the next 'charged' tRNA to the A site. If the anticodon of the tRNA cannot pair with the A site, then the tRNA is released and this process is repeated until the correct 'charged' tRNA is inserted into the A site. Meanwhile, the deacylated tRNA in the E site abandons the ribosome. Next, a new peptide bond between the amino acid carried by the tRNA in the P site and the amino acid carried by the tRNA in the A site is formed, leaving a free tRNA in the P site and a peptidyl tRNA in the A site. The last step of the elongation is translocation. During translocation, the ribosome moves one codon forward, making the free tRNA in the P site transferred to the E site and the peptidyl tRNA in the A site transferred to the P site. Thereafter, the elongation process goes back to the first step: the E site is bound by the free tRNA, the P site is bound by the peptidyl tRNA and the A site is empty. As Figure 1.2 illustrates, the ribosome is ready for binding of a new 'charged' tRNA to the A site and incorporating a new amino acid to the growing polypeptide.

The elongation process is iterated until the A site of the ribosome encounters a stop codon. In the cell, there is no tRNA with anticodon complementary to the stop codon. Rather, cells contain certain *release factors* that respond to the termination signals to terminate the translation process. The release factors bind to the stop codon at the A site and hydrolyse the bond between the polypeptide chain and the tRNA bound in the P site, releasing a newly synthesised protein. After that, the ribosome subunits dissociate from the mRNA.

1.1.2 mRNA Degradation

Stability of mRNA is important to the translation. The greater the mRNA stability, the more protein can be produced from that mRNA. Lifetimes of prokaryotic mRNA can be in the range of seconds to over an hour [117, 141]. While eukaryotic mRNA lifetimes range from minutes to days [149]. The mechanisms for mRNA degradation are different in prokaryotes and eukaryotes.

In general, the prokaryotic mRNA degradation depends on a rate limiting initiation step, following first order kinetics [38, 92]. After initiation, endonucleases cleavage the mRNA into fragments in the 5' end to the 3' end direction. Subsequently, those fragments are digested to nucleotides by exonucleases in the 3' end to the 5' end direction [94].

Eukaryotic mRNA degradation is determined by the mRNA structure, namely, *poly(A) tail* and *5' cap*. The poly(A) tail consists of multiple adenosine monophosphates in the 3' end of an mRNA. The 5' cap is a specially altered nucleotide on the 5' end of an mRNA [60]. In general, eukaryotic mRNA degradation requires three steps [94, 131]. The first step is deadenylation, i.e., removal of the poly(A) tail. The second step is decapping: hydrolysis of the 5' cap structure on an mRNA. The last step is to degrade

the mRNA by a combination of endonucleases and exonucleases.

Other than the above general degradation pathways, mRNA degradation is also mediated by other mechanisms. For instance, *nonsense mediated decay* to reduce errors in gene expression by eliminating mRNA that contain premature stop codons [6]; microRNAs can accelerate the deadenylation, thereby hastening mRNA degradation [145]; small interfering RNA (siRNA) mediated mRNA breaking down [112].

1.1.3 Regulation of Translation

Regulations can occur throughout the whole translation process: initiation, elongation and termination [55]. However, in most cases, initiation is a key and rate limiting step. Thus, translation regulation primarily targets the initiation stage [89, 102]. The mechanisms of translational regulation are different for eukaryotes and prokaryotes for a number of reasons, for instance, (1) prokaryotic translation begins while the mRNA is still being synthesised, however, eukaryotic mRNA has to be transported to the cytoplasm before translation [59, 105]; (2) compared to eukaryotic translation, there are relatively large number of ribosomes translating the same mRNA, thus, ribosomes do not have to be recycled like in eukaryotes (we will introduce this recycling later) [61]; (3) short mRNA lifetime is sufficient to supply protein for a small and quickly dividing prokaryotic cell, while eukaryotic cells require more stable mRNAs with 5' cap, poly(A) tail or other secondary structures.

In prokaryotes, there are three main translational regulation pathways: through trans-acting proteins, through trans-acting RNAs, and through cis-acting mRNA elements. **(I) Protein mediated regulation**. The majority of proteins involved in this regulation pathway act as repressors. A typical example is ribosomal protein S1 in *E. coli*. S1 participates in translation initiation complex formation by assisting the positioning of the

small ribosome subunit in the translation initiation region. However, redundant S1 can bind to the AU rich region of its own ribosome binding site, inhibiting its own translation initiation [40, 116]. **(II) RNA mediated regulation.** *Antisense RNAs (asRNAs)* can bind to the target mRNA to curtail initiation and even lead to mRNA degradation [139]. asRNA is a single-stranded RNA that is complementary to an mRNA. Bacteria *small RNAs (sRNAs)* can bind to mRNA targets and regulate translation. sRNAs are small (50-250 nucleotide) non-coding RNA molecules. For example, DsrA activates translation of the stress sigma factor, RpoS in *E. coli* [101]. **(III) cis-mRNA elements mediated regulation.** mRNA can even function as proteins or sRNAs to regulate its translation. For instance, the *prfA* mRNA in *L. monocytogenes* has a secondary structure in 5' untranslated region to inhibit translation at low temperature (below 37°) [84]. Another cis-mRNA structure is *riboswitch*, a regulatory segment that binds a small molecule, resulting in repression or activation of its own translation [9, 111].

For eukaryotes, we introduce two main regulation categories. **(I) Regulation by RNA-binding proteins.** Some proteins are able to interact with the 5' untranslated region of an mRNA to inhibit translation, e.g., poly(A) binding protein, or PABP, can bind to clustered motifs that are a few nucleotides downstream of the 5' cap of a *PABP* mRNA, blocking ribosome scanning [110]. Important protein-mRNA interaction occurs between PABP and the 3' poly(A) tail of an mRNA. PABP are able to bind to the poly(A) tail and further be bound by the 5' end of the mRNA, resulting in a circularisation configuration ('closed loop') to allow ribosome recycling and translation reinitiation [3]. However, the 'closed loop' may also be constructed to inhibit translation. For example, in *Drosophila melanogaster* embryos, protein Smaug binds to the 3' untranslated region of *Nanos* mRNA, interacting with eIF4E1a to form an inhibitory 'closed loop' with the help of bridging protein CUP [128]. **(II) Regulation by microRNAs.** MicroRNA is a small non-coding RNA molecule found in eukaryotes, that is complementary to a part of one or more mRNAs. MicroRNA negatively

Component	Dry Cell Mass (%)	Molecules /Cell	Different types	Copies of each type
Wall	10	1	1	1
Membrane	10	2	2	1
DNA	1.5	1	1	1
mRNA	1	1,500	600	$2-3$
tRNA	3	200,000	60	>3000
rRNA	16	38,000	2	19,000
Ribosomal protein	9	10^6	52	19,000
Soluble protein	46	2.0×10^6	1850	>1000
Small molecules	3	7.5×10^6	800	

Table 1.1: **Considering *E. coli* in terms of its macromolecular components.** 25% of bacterial dry mass is concerned with gene expression. Table is reproduced from [94].

regulates translation, the mechanism including: preventing assembly of the initiation complex or recruiting the ribosomal subunit in the initiation stage; inhibition of the elongation; responsible for ribosome drop off (translation premature termination) and mRNA degradation [5, 71, 107, 114, 140].

1.1.4 Stochasticity in Translation

Experimental data have demonstrated that gene expression is a fundamentally stochastic process [18, 52]. All cellular events depend on the stochastic molecular collisions. Mathematically, if numerous identical events occur independently, then stochastic fluctuations can be ignored. However, in practice, active mRNA molecules exist in a low number of copies in a cell, making stochasticity important for the translation process. Table 1.1 presents data on the macromolecular components for 25% of *E. coli* dry mass. Typically, only 2-3 copies of particular mRNA are found. Moreover, the translation process relies on several biochemical reactions such as ribosome assembly, tRNA pairing and other molecular regulations. These random association and disassociation events would undoubtedly introduce considerable noise. The stochasticity motivates stochastic modelling for the translation process, and we will introduce it later.

1.2 Mathematical Background

Mathematical modelling of the translation process has attracted comprehensive attention. 'Bottom up' (understand the process from known, simple behaviour of its components) and 'top down' (understand the process by breaking down it compositional subsystems) approaches have been employed to study the translation process [36, 126, 133, 138]. Two broad classes of mathematical modelling have been developed: deterministic approach and stochastic approach. We now discuss each approach in turn.

1.2.1 Deterministic Approach

Determinstic approaches are perhaps most commonly used to model chemical reactions [7]. Systems of ordinary differential equations (ODEs) and delay differential equations (DDEs) are the essential tools of the determintic approach, namely, using *rate equation* to express the rate of production of a component of the system as a function of the concentrations of other components:

$$\begin{cases} \text{ODE:} & \dfrac{dx_i(t)}{dt} = f_i\left(\boldsymbol{x(t)}\right), \qquad i \leqslant i \leqslant n, \\ \text{DDE:} & \dfrac{dx_i(t)}{dt} = f_i\left(\boldsymbol{x(t-\tau_i)}\right), \qquad i \leqslant i \leqslant n, \end{cases} \tag{1.2.1}$$

where $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$ is the concentration of all the components of the system, τ_i is time delay and f_i is usually a nonlinear function [35, 51].

The translation process involves a series of biochemical reactions: ribosome assembly, reading the mRNA codon by codon, detaching the mRNA, releasing a protein and all the other possible regulations. However, it is not necessary to model every aspect of the whole process. Translation initiation is a key rate limiting step. Therefore, using

ODE, the simplest chemical reaction system to describe the translation process is as follows (the elongation detail is omitted) [142]:

$$\begin{aligned} \text{mRNA} + \text{ribosome} &\longrightarrow \text{mRNA}\cdot\text{ribosome}, \\ \text{mRNA}\cdot\text{ribosome} &\longrightarrow \text{mRNA} + \text{ribosome} + \text{protein}. \end{aligned}$$

In this simplest case, only three ODEs are sufficient to describe the translation process. To take elongation into account, the translation process is typically modelled as a set of ODEs by considering the following chemical reaction systems [50, 75, 127, 154]:

$$\begin{aligned} \text{RBS} + \text{ribosome} &\longrightarrow \text{RBS}\cdot\text{ribosome}, \\ \text{RBS}\cdot\text{ribosome} &\longrightarrow \text{RBS}\cdot\text{ribosome}\cdot\text{P}_1, \\ \text{RBS}\cdot\text{ribosome}\cdot\text{P}_1 &\longrightarrow \text{codon}_2\cdot\text{ribosome}\cdot\text{P}_2, \\ \vdots &\longrightarrow \vdots \\ \text{codon}_{L-1}\cdot\text{ribosome}\cdot\text{P}_{L-1} &\longrightarrow \text{codon}_L\cdot\text{ribosome}\cdot\text{P}_L, \\ \text{codon}_L\cdot\text{ribosome}\cdot\text{P}_L &\longrightarrow \text{ribosome} + \text{mRNA} + \text{protein}, \end{aligned}$$

where P_i denote a polypeptide with length i. In the above system, the first biochemical reaction describes translation initiation, the last biochemical reaction indicates translation termination and biochemical reactions inbetween characterise translation elongation. Each possible codon·ribosome·polypeptide complex is regarded as an independent biochemical species. Thus, $L+2$ ODEs are required to describe this system (L ODEs for all the possible biochemical species, plus two extra ODEs for initiation and termination). Clearly then, if initiation, termination and elongation steps are presented by a series of biochemical reactions, rather than a single simple step, the resultant ODE system can be much more complicated [16, 37, 45, 153].

More recently, to capture the fact that there exist time delays in the translation process, delay differential equation (DDE) models have been developed [83, 119, 150, 151]. In those works, the translation process is basically described by:

$$\text{ribosome}(t)+\text{RBS}(t)\longrightarrow \text{RBS}(t+\tau_1)+\text{ribosome}(t+\tau_2)+\text{protein}(t+\tau_2). \quad (1.2.2)$$

It is assumed that a ribosome is bound to the RBS for a time τ_1. The ribosome then transits the mRNA in a time τ_2 ($\tau_2 > \tau_1$). At this stage, the ribosome detaches from the mRNA and a protein is produced.

The above deterministic approach does not reflect some features of the translation process, for example, the fact that a codon cannot be simultaneously occupied by more than one ribosome. Moreover, considering the large size of the ODE system that may be involved, analytical solutions to the ODE system are difficult to obtain.

1.2.2 Stochastic Approach

An alternative approach to model the translation process is stochastic modelling using the so called *TASEP*, the totally asymmetric exclusion process, which was first introduced in the 1960s [98, 99, 152].

The schematic of the TASEP is illustrated in Figure 1.3. In this model, mRNA is represented by a one dimensional lattice. Codons on the mRNA are represented by sites labelled by $i = 1, 2, \ldots, L$. Ribosomes are referred to as particles. For simplicity, the size of a particle is assumed to be equivalent to the size of a site. Site i can be occupied by 0 or 1 particle (described by $S_i = 0$ or 1), so that the system configuration can be described by a state vector $(S_1, S_2, \ldots, S_L)$. Three major assumptions are made: (i) particles move unidirectionally, the size of a particle is the same as that of a site and the supply of particles is unlimited; (ii) motion between neighbouring particles is

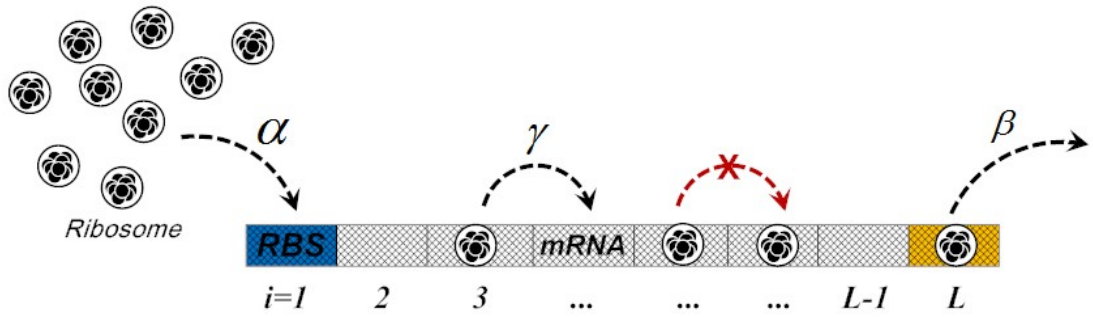


Figure 1.3: **Schematic of the TASEP.** Particles move unidirectionally with rate γ. No more than one particle can occupy a site simultaneously. Entry rate is denoted by α and exit rate is given by β.

steric exclusive, which ensures no more than one particle can simultaneously occupy one site; (iii) a single, uniform elongation rate is applied. To capture the essence of the translation process, the particle-hole exchange rule of the TASEP is as follows. Particles bind to site 1 at rate α provided $S_1 = 0$ (corresponding to the translation initiation); a particle translocation takes place between site i and $i+1$ at rate γ provided $S_i = 1, S_{i+1} = 0$ (corresponding to the elongation step, and γ is usually rescaled to be 1); the particle on site L detaches from the lattice at rate β (corresponding to the translation termination). Let ρ_i be *density profile*, which is the average occupation of site i in the long time limit. Let ρ denote *overall average density*, and it is given by $\rho = \sum \rho_i / L$. The *current*, J, quantifying the average number of particles that enter/exist the lattice, is defined as:

$$J = \langle S_i (1 - S_{i+1}) \rangle = \alpha \langle 1 - S_1 \rangle = \beta \langle S_L \rangle . \tag{1.2.3}$$

Though the evolution of the TASEP is hard to predict, a steady state (time independent) solution of ρ_i, ρ, and J can be obtained. For system with large size L, mean field approximation is applied to solve the steady state. The essence of the mean field approximation is to ignore the interplay between the neighbouring particles, namely,

$$J = \rho_i (1 - \rho_{i+1}) = \alpha (1 - \rho_1) = \beta \rho_L . \tag{1.2.4}$$

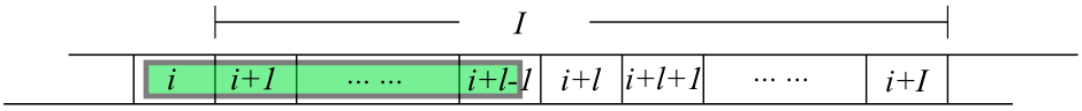


Figure 1.4: **TASEP with extended particle size $l > 1$.** Consider a large section of length I in the lattice. A ribosome (green object) with size l locates its reader at site i.

Moreover, the relationship between the overall average density ρ and the current J is given by:

$$J = \rho(1-\rho). \tag{1.2.5}$$

The exact and mean field steady state solution has been solved [43, 44]. We will reproduce these results and fill in gaps in Chapter 2.

To make the proto TASEP model more realistic, Gibbs *et al.* [98, 99] took into account the large size of ribosomes compared to a codon, i.e., considered TASEP with particle size $l > 1$. Experimental data suggest that $l = 12$ is the most appropriate [75]. In this generalised TASEP, a particle occupies l sites on the lattice, i.e., a functional particle at site i is extended to occupy sites i to site $i+l-1$ and site i is referred to as a 'reader' [124]. The particle-hole exchange rule is modified. To enter the lattice, the first l sites must be empty. The particle with the 'reader' on site i translocates if site $i+l$ is empty. While a particle is reading the last l sites, there is no other particle forward, thus there is no steric hindrance for the last l hops and the particle leaves the lattice at the site L (the 'reader' of the large sized particle is on site L) [93]. In this case, the current J is given by:

$$J = \alpha \left\langle \prod_{k=1}^{l} (1-S_k) \right\rangle = \left\langle S_i (1-S_{i+l}) \right\rangle = \langle S_{L-l} \rangle = \ldots = \langle S_{L-1} \rangle = \beta \langle S_L \rangle . \tag{1.2.6}$$

Consider a fixed section of length I in the lattice (I is large), as Figure 1.4 shows. Suppose the ribosome in Figure 1.4 is the jth particle from the left end of the lattice, and the reader is denoted by $x_j = i$. Then the probability that the ribosome is free to

translocate (i.e., $S_{i+l} = 0$) is given by:

$$P\left(x_{j+1} > i+l\right) = \frac{S_{i+l+1} + S_{i+l+2} + \ldots + S_{i+I}}{S_{i+l} + S_{i+l+1} + \ldots + S_{i+I}} = \frac{\sum_{k=i+1}^{I} S_k - \sum_{k=i+1}^{i+l} S_k}{\sum_{k=i+1}^{I} S_k - \sum_{k=i+1}^{i+l} S_k + S_{i+l}}$$
$$= \frac{1 - \sum_{k=i+1}^{i+l} \rho_k}{1 - \sum_{k=i+1}^{i+l} \rho_k + \rho_{i+l}} = \frac{\bar{\rho}_i}{\rho_{i+l} + \bar{\rho}_i}, \qquad \text{where} \quad \bar{\rho}_i = 1 - \sum_{k=i+1}^{i+l} \rho_k,$$

which is also the probability that the location of reader of the $j+1$th particle satisfies $x_{j+1} > i+l$. Thus, the current can be calculated by $\rho_i P\left(x_{j+1} > i+l\right)$ as:

$$J = \frac{\rho_i \bar{\rho}_i}{\rho_{i+l} + \bar{\rho}_i}. \tag{1.2.7}$$

To better predict the density profile, Shaw *et al.* [123] considered pair correlations between particles. However, no higher level of mean field theory has been developed.

Another important modification to the standard TASEP is to consider the site dependent hopping rates. In practice, the hopping rates depend on the nucleotide sequence and the abundance of the specific 'charged' tRNA, resulting in an inhomogeneous hopping rate sequence γ_i. Such modification makes it extremely difficult to determine the density profile, overall average density and current. Most progress was made mainly with numerical simulations while some understanding is established by using mean field approximation [47, 56, 87]. With a full set of random hopping rates γ_i which are chosen from some known distributions, research has concentrated on the effect of the disorder on the relationship between the overall average density and the current [11, 68, 74, 90, 132]. In particular, *slow site* research has attracted broad attention. The slow site is a site has slower hopping rate than all its neighbours. The density profile and current are affected by the location of the slow site. If there are two slow sites, the distance between them dramatically affects the current [25, 46, 56]. For example, Romano *et al.* [121] showed that the queueing of the particles behind a slow site can lead to a first-order phase transition. However, all these works were done for $l = 1$,

inhomogeneous TASEP with $l > 1$ remain to be uncovered.

Apart from extended particle size and inhomogeneous hopping rates, realistic translation model also requires limited supply of resources, for example, two most important components: ribosome and tRNA. A modification has been made to allow the particle binding rate to mediate the available ribosomes in the reservoir [1, 29, 30]. In those works, a single TASEP was studied and the total number of particles was fixed and denoted by a constant N_{tot}. The effective particle binding rate was assumed to be:

$$\alpha_{eff} = \alpha \tanh\left(\frac{n_0}{N^*}\right), \tag{1.2.8}$$

where n_0 was the particles in the reservoir and N^* was a crossover parameter. Let N denote the particles on the lattice, then $N + n_0 = N_{tot}$. Results showed that the overall average density and the current were affected by N_{tot} and the length of the lattice, L. Competition for the finite particle resource between multiple TASEPs has also been studied [31, 67]. Domain Wall theory [31] and mean field approximation [67] were adopted respectively to capture the nontrivial behaviours of the multiple TASEPs with different length, particle binding/exit rate. Results showed that if all TASEPs have equal length, then the behaviour is analogous to that of a single TASEP with finite particle supply. For the case that involved different lattice lengths, several unanticipated new regimes emerged. Brackley *et al.* [19] considered the finite supply of tRNA. In their work, the effective hopping rate was mediated by the available 'charged' tRNA in the reservoir:

$$\gamma_{eff} = \gamma T, \tag{1.2.9}$$

where T denotes the 'charged' tRNA and γ is the intrinsic elongation rate. The total number of tRNA was fixed as a constant, $\bar{T}$. The formation of the 'charged' tRNA was assumed to follow Michaelis-Menten kinetics. Results showed that limiting the supply of tRNA dramatically reduced the current.

More complexity has been included in the modelling of the elongation step [8, 27, 62, 63, 77, 118]. In those works, the elongation step consisted of several sub steps, rather than being described by a single, random hopping event. Usually, the effective hopping rate is derived first, making it possible to apply the framework of the standard TASEP. Shifted phase diagrams are obtained due to the varying effective hopping rate. For example, Ciandrini *et al.* [27] suggested that the elongation step could be approximated by two sub steps: (i) searching for the correct 'charged' tRNA; (ii) translocation from one site to the next. Thus a ribosome was represented by a two-state particle: state 1 denoted the absence of the correct 'charged' tRNA while state 2 denoted the presence. The transition rate from the state 1 to state 2 was k. The translocation can only occur at rate γ if a particle was in state 2.

More complex translation initiation and termination have been considered [17, 37, 45, 147]. One important example is that eukaryotic mRNA can form a 'closed loop' due to the interaction between the poly(A) tail and PABP. This enhances the efficiency of translation. Two main hypotheses have been taken to understand this translation efficiency enhancement. One hypothesis is that the efficiency enhancement is due to the ribosome diffusion [24, 64, 103, 122]. In those models, ribosomes were assumed to leave the 3' end of the mRNA to enter the ribosome reservoir. Due to the 'closed loop', the local concentration of ribosomes around the ribosome binding site (located in the 5' end of the mRNA) was increased. Thus, the ribosome binding rate is increased. The other hypothesis is motivated by the experimental evidence that the enhancement of efficiency of the translation is not only due to the passive ribosome diffusion. Ribosomes can pass from the 3' end to the 5' end of the mRNA directly [48, 57, 58, 115, 125, 146]. Marshall *et al.* [104] developed a ribosome recycling model based on TASEP to incorporate ribosome reinitiation. Essentially, Marshall *et al.* [104] assumed that ribosomes could be recruited directly from the stop codon to the start codon to initiate translation.

Computational methods are used to simulate TASEP based model. In general, there are

two main methods. One is the Gillespie Stochastic Simulation Algorithm, or Gillespie SSA [65, 66, 78, 79]. In the Gillespie SSA, randomly generated numbers are used to (i) determine which reaction will occur, and (ii) when the chosen reaction occurs. The other technique is the Monte Carlo method [95, 136]. In the Monte Carlo method, randomly generated numbers are used to (i) choose one reaction from all the possible reactions, and (ii) determine whether or not the chosen reaction occurs. Details will be presented later.

1.3 Thesis Outline

The rest of this thesis is ordered as follows. In Chapter 2 we review key properties of the TASEP and reproduce the exact and mean field solution of the steady state. We fill some gaps in the steps presented in the original works [43, 44]. Next, we develop a linear approximation to measure the transient state of the TASEP. Particularly, we approximate the rate of particles enter/exit the lattice before the steady state is achieved. Finally, we give a introduction to the Gillespie SSA and the Monte Carlo methods.

In Chapter 3 we investigate the role of mRNA degradation in mediating translation. We assume that mRNA degradation is initiated by a simple, single step. After initiation, (i) endonucleases track down the last ribosome on the mRNA in 5'$-$3' direction to degrade the mRNA, or (ii) endonucleases digest the mRNA quickly. We report that the simple single step of mRNA degradation initiation can be modelled as a Poisson process. We attempt to derive the probability density function of protein production during the whole mRNA lifetime. We also investigate how the parameters affect the probability density function. Differences between the two mRNA degradation types are discussed.

In Chapter 4, we consider a general case of negative feedback control on mRNA translation initiation. It is assumed that binding of ribosomes to the RBS of a given mRNA is suppressed by proteins produced from that mRNA itself. We extend this model in Chapter 5 to include ribosome recycling. In Chapter 4 and Chapter 5, we attempt to derive the effective ribosome binding/exit rate in order to apply the framework of the mean field solution of the TASEP. We obtain the generalised phase diagram, as well as the average density and the current in all phases. We investigate how the auto-negative feedback/auto-negative feedback and ribosome recycling affects the protein level. DDE model is developed to analyse the dynamics. Special cases of oscillation and bistable switching are found.

In Chapter 6 we summarise key results and discuss the possible implications of obtaining a better understanding of the translation process. Possible future directions for the application of the TASEP technique and the modelling of the translation are discussed.

Chapter 2

TASEP and its Simulation

In this chapter, we first follow the works of Derrida *et al.* [43, 44] to derive an exact solution and also a mean field solution for TASEP in long time limit. The proofs in [43, 44] omit many steps. We summarise the results in [43, 44] as lemmas and fill the steps here to provide the reader with a more accessible description of the arguments. We define phases of TASEP in a more rigorous way. Next, we give an approximation of transient state of TASEP before the steady state is reached. Finally we introduce two main simulation methods, Gillespie Stochastic Simulation Algorithm and Monte Carlo simulation method. This chapter is the basis of the whole thesis.

2.1 Exact Steady State Solution

TASEP is a stochastic process, particles enter into the lattice from the left end and leave the lattice from the right end with fixed rates. Let $P_L(S_1, S_2, \ldots, S_L)$ be the probability that TASEP is in the state $(S_1, S_2, \ldots, S_L)$. $P_L(S_1, S_2, \ldots, S_L)$ can be scaled by a constant

Z_L as:

$$P_L(S_1, S_2, \ldots, S_L) = \frac{f_L(S_1, S_2, \ldots, S_L)}{Z_L}, \tag{2.1.1}$$

where

$$Z_L = \sum_{S_1=0,1} \cdots \sum_{S_L=0,1} f_L(S_1, S_2, \ldots, S_L).$$

Derrida *et al.* [43] argued that it would be more convenient to use $f_L(S_1, S_2, \ldots, S_L)$ afterwards. It is shown that in steady state, $f_L(S_1, S_2, \ldots, S_L)$ is simply given by a product of L matrices D or E with matrix D at site i if site i is occupied and matrix E if site i is vacant[43, 53, 72, 86]:

$$f_L(S_1, S_2, \ldots, S_L) = \langle W | \prod_{i=1}^{L} \left(S_i D + (1 - S_i) E \right) | V \rangle \tag{2.1.2}$$

where D, E are square matrices and $\langle W|$, $|V\rangle$ are vectors. In general, it is not necessary to specify these matrices and vectors. Moreover,

Lemma 2.1. *TASEP is in steady state if:*

$$DE = D + E, \qquad D|V\rangle = \frac{1}{\beta}|V\rangle, \qquad \langle W|E = \frac{1}{\alpha}\langle W|. \tag{2.1.3}$$

Proof. (see [44]) To prove Lemma 2.1, there is need to introduce transition rate matrix. The transition rate matrix Q is an $n \times n$ matrix whose elements describe the rates a stochastic process moves among the n states. The elements q_{ij}, $i \neq j$, are non-negative and represent the rates that the process moves between the state i and j. The diagonal elements q_{ii} are chosen such that each row of Q sums to zero, i.e., $q_{ii} = -\sum_{j \neq i} q_{ij}$. q_{ii} represent the rates that a stochastic process out of state i.

Consider the transition rate matrix for particle entering the lattice from the left end at rate α. There are two states for this process, $(0, S_2, \ldots, S_L)$ and $(1, S_2, \ldots, S_L)$. Let h_1 denote this transition rate matrix, the size of h_1 is 2×2. The transition rate from state

$(0, S_2, \ldots, S_L)$ to $(1, S_2, \ldots, S_L)$ is represented by $(h_1)_{1;0}$, which is given by α. The transition rate out of the state $(0, S_2, \ldots, S_L)$ is $(h_1)_{0;0}$. $(h_1)_{0;0}$ is chosen to be $-\alpha$ such that the first row of h_1 sums to zero. In TASEP, particles move unidirectionally from the left end to the right end of the lattice, so the transition rate from state $(1, S_2, \ldots, S_L)$ to the state $(0, S_2, \ldots, S_L)$ and the transition rate out of state $(1, S_2, \ldots, S_L)$ are both 0. The transition rate matrix is therefore given by:

$$h_1 = \begin{pmatrix} -\alpha & \alpha \\ 0 & 0 \end{pmatrix}.$$

Consider the transition rate matrix for particle leaving from the right end of the lattice at rate β. There are two states for this process, $(S_1, S_2, \ldots, 0)$ and $(S_1, S_2, \ldots, 1)$. Let h_L represent this matrix. The size of h_L is 2×2. $(h_L)_{0;1}$, the transition rate from the state $(S_1, S_2, \ldots, 1)$ to the state $(S_1, S_2, \ldots, 0)$, is given by β. The transition rate out of the state $(S_1, S_2, \ldots, 1)$, $(h_L)_{1;1}$, is chosen to be $-\beta$ such that the second row of h_L sums to zero. The remaining two elements of h_L are both 0 because particles cannot enter into the lattice from the right end. The transition rate matrix is therefore given by:

$$h_L = \begin{pmatrix} 0 & 0 \\ \beta & -\beta \end{pmatrix}.$$

Consider the transition rate matrix for particles hopping between the non-boundary site i and $i+1$, $1 \leqslant i < L$. There are four states for this process, $(\ldots, S_{i-1}, 0, 0, S_{i+1}, \ldots)$, $(\ldots, S_{i-1}, 0, 1, S_{i+1}, \ldots)$, $(\ldots, S_{i-1}, 1, 0, S_{i+1}, \ldots)$ and $(\ldots, S_{i-1}, 1, 1, S_{i+1}, \ldots)$. Let h denote this matrix, the size of h is 4×4. The only transition that can occur is the transition from the state $(\ldots, S_{i-1}, 1, 0, S_{i+1}, \ldots)$ to the state $(\ldots, S_{i-1}, 0, 1, S_{i+1}, \ldots)$. The transition rate for this transition, $(h)_{0,1;1,0}$, is 1 (note that γ is set to be 1). The transition rate out of the state $(\ldots, S_{i-1}, 1, 0, S_{i+1}, \ldots)$, $(h)_{1,0;1,0}$, is therefore chosen

to be -1, such that the row sums to zero. All the other transition rates are 0. The transition rate matrix is then given by:

$$h = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

In summary, the non-zero elements of these transition rate matrices are

$$\begin{aligned} (h_1)_{0;0} &= -\alpha & (h_1)_{1;0} &= \alpha \\ (h)_{1,0;1,0} &= -1 & (h)_{0,1;1,0} &= 1 \\ (h_L)_{1;1} &= -\beta & (h_L)_{0;1} &= \beta. \end{aligned} \tag{2.1.4}$$

The evolution of $P_L(S_1, S_2, \ldots, S_L)$ can be described by the following master equation, using the transition rate matrices,

$$\begin{aligned} \frac{d}{dt} P_L(S_1, S_2, \ldots, S_L) = &\sum_{\tau_1} (h_1)_{S_1;\tau_1} P_L(\tau_1, S_2, \ldots, S_L) \\ &+ \sum_{i=1}^{L-1} \sum_{\tau_i, \tau_{i+1}} (h)_{S_i, S_{i+1}; \tau_i, \tau_{i+1}} P_L(S_1, \ldots, \tau_i, \tau_{i+1}, \ldots, S_L) \\ &+ \sum_{\tau_L} (h_L)_{S_L;\tau_L} P_L(S_1, \ldots, S_{L-1}, \tau_L), \end{aligned} \tag{2.1.5}$$

where $\tau_i \in \{0, 1\}$ is the possible occupation of site i. We are principally interested in steady state of (2.1.5), that is conditions which ensure the right hand side of (2.1.5) is equal to 0. If there exist two coefficients x_0, x_1 satisfying the following equalities:

$$\sum_{\tau_1} (h_1)_{S_1;\tau_1} f_L(\tau_1, S_2, \ldots, S_L) = x_{S_1} f_{L-1}(S_2, \ldots, S_L), \tag{2.1.6}$$

$$\sum_{\tau_i,\,\tau_{i+1}} (h)_{S_i,S_{i+1};\,\tau_i,\,\tau_{i+1}} f_L(S_1,\ldots,S_{i-1},\tau_i,\tau_{i+1},S_{i+2},\ldots,S_L)$$

$$= -x_{S_i} f_{L-1}(S_1,\ldots,S_{i-1},S_{i+1},\ldots,S_L)$$

$$+ x_{S_{i+1}} f_{L-1}(S_1,\ldots,S_i,S_{i+2},\ldots,S_L), \tag{2.1.7}$$

$$\sum_{\tau_L} (h_L)_{S_L;\,\tau_L} f_L(S_1,\ldots,S_{L-1},\tau_L) = -x_{S_L} f_{L-1}(S_1,\ldots,S_{L-1}). \tag{2.1.8}$$

then indeed TASEP is in steady state since by substituting (2.1.6-2.1.8) into (2.1.5) yields $\dfrac{dP_L(S_1,S_2,\ldots,S_L)}{dt} = 0$. Now the task is to find the coefficients x_0, x_1 using (2.1.2, 2.1.6-2.1.8).

For (2.1.6), if $S_1 = 0$, the only non-zero transition rate is $(h_1)_{0;0}$, therefore:

$$\sum_{\tau_1} (h_1)_{S_1;\,\tau_1} f_L(\tau_1,S_2,\ldots,S_L)$$

$$= \sum_{\tau_1} (h_1)_{0;\,\tau_1} f_L(\tau_1,S_2,\ldots,S_L) = (h_1)_{0;0} f_L(0,S_2,\ldots,S_L)$$

$$= -\alpha \langle W|E \prod_{i=2}^{L} \big(S_i D + (1-S_i)E\big)|V\rangle,$$

$$x_{S_1} f_{L-1}(S_2,\ldots,S_L) = x_0 \langle W| \prod_{i=2}^{L} \big(S_i D + (1-S_i)E\big)|V\rangle,$$

therefore

$$-\alpha \langle W|E \prod_{i=2}^{L} \big(S_i D + (1-S_i)E\big)|V\rangle = x_0 \langle W| \prod_{i=2}^{L} \big(S_i D + (1-S_i)E\big)|V\rangle,$$

then

$$\alpha \langle W|E = -x_0 \langle W|.$$

If $S_1 = 1$, the only non-zero transition rate is $(h_1)_{1;0}$, therefore

$$\sum_{\tau_1} (h_1)_{S_1;\,\tau_1} f_L(\tau_1,S_2,\ldots,S_L)$$

$$
\begin{aligned}
&= \sum_{\tau_1} (h_1)_{1;\tau_1} f_L(\tau_1, S_2, \ldots, S_L) = (h_1)_{1;0} f_L(0, S_2, \ldots, S_L) \\
&= \alpha \langle W | E \prod_{j=2}^{L} \left(S_i D + (1 - S_i) E \right) |V\rangle, \\
&\quad x_{S_1} f_{L-1}(S_2, \ldots, S_L) = x_1 \langle W | \prod_{j=2}^{L} \left(S_i D + (1 - S_i) E \right) |V\rangle, \\
&\quad \therefore \alpha \langle W | E = x_1 \langle W |.
\end{aligned}
$$

Overall,

$$
\alpha \langle W | E = x_1 \langle W | = -x_0 \langle W |. \tag{2.1.9}
$$

For (2.1.7), if $S_i = 0$, the only non-zero transition rate is $(h)_{0,1;1,0}$, therefore:

$$
\begin{aligned}
&\sum_{\tau_i, \tau_{i+1}} (h)_{S_i, S_{i+1}; \tau_i, \tau_{i+1}} f_L(S_1, \ldots, S_{i-1}, \tau_i, \tau_{i+1}, S_{i+2}, \ldots, S_L) \\
&= (h)_{0,1;1,0} f_L(S_1, \ldots, S_{i-1}, 1, 0, S_{i+2}, \ldots, S_L) \\
&= \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) DE \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle, \\
&\quad - x_{S_i} f_{L-1}(S_1, \ldots, S_{i-1}, S_{i+1}, \ldots, S_L) + x_{S_{i+1}} f_{L-1}(S_1, \ldots, S_i, S_{i+2}, \ldots, S_L) \\
&= - x_0 f_{L-1}(S_1, \ldots, S_{i-1}, 1, \ldots, S_L) + x_1 f_{L-1}(S_1, \ldots, 0, S_{i+2}, \ldots, S_L) \\
&= - x_0 \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) D \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle \\
&\quad + x_1 \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) E \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle,
\end{aligned}
$$

$$
\begin{aligned}
&\therefore \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) DE \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle \\
&= - x_0 \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) D \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle \\
&\quad + x_1 \langle W | \prod_{j=1}^{i-1} \left(S_j D + (1 - S_j) E \right) E \prod_{j=i+2}^{L} \left(S_j D + (1 - S_j) E \right) |V\rangle,
\end{aligned}
$$

then

$$DE = -x_0 D + x_1 E.$$

If $S_i = 1$, the only non-zero transition rate is $(h)_{1,0;1,0}$, therefore:

$$\begin{aligned}
&\sum_{\tau_i,\,\tau_{i+1}} (h)_{S_i,S_{i+1};\tau_i,\tau_{i+1}} f_L(S_1,\ldots,S_{i-1},\tau_i,\tau_{i+1},S_{i+2},\ldots,S_L) \\
&= (h)_{1,0;1,0} f_L(S_1,\ldots,S_{i-1},1,0,S_{i+2},\ldots,S_L) \\
&= -\langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) DE \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle, \\
&\quad - x_{S_i} f_{L-1}(S_1,\ldots,S_{i-1},S_{i+1},\ldots,S_L) + x_{S_{i+1}} f_{L-1}(S_1,\ldots,S_i,S_{i+2},\ldots,S_L) \\
&= -x_1 f_{L-1}(S_1,\ldots,S_{i-1},0,\ldots,S_L) + x_0 f_{L-1}(S_1,\ldots,1,S_{i+2},\ldots,S_L) \\
&= -x_1 \langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) E \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle \\
&\quad + x_0 \langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) D \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle, \\
&\therefore -\langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) DE \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle \\
&= -x_1 \langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) E \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle \\
&\quad + x_0 \langle W| \prod_{j=1}^{i-1} \Big(S_j D + (1-S_j) E\Big) D \prod_{j=i+2}^{L} \Big(S_j D + (1-S_j) E\Big) |V\rangle,
\end{aligned}$$

then

$$DE = -x_0 D + x_1 E.$$

Overall, one gets

$$DE = -x_0 D + x_1 E. \tag{2.1.10}$$

For (2.1.8), if $S_L = 0$, the only non-zero transition rate is $(h_L)_{0;1}$, therefore

$$\sum_{\tau_L} (h_L)_{S_L;\tau_L} f_L(S_1,\ldots,S_{L-1},\tau_L)$$

$$
\begin{aligned}
&= \sum_{\tau_L} (h_L)_{0;\,\tau_L} f_L(S_1,\ldots,S_{L-1},\tau_L) = (h_L)_{0;1} f_L(S_1,\ldots,S_{L-1},1) \\
&= \beta \langle W| \prod_{j=1}^{L-1} \Big(S_j D + \big(1-S_j\big) E \Big) D|V\rangle, \\
&\quad - x_{S_L} f_{L-1}(S_1,\ldots,S_{L-1}) = -x_0 \langle W| \prod_{j=1}^{L-1} \Big(S_j D + \big(1-S_j\big) E \Big) |V\rangle,
\end{aligned}
$$

then

$$\beta D|V\rangle = -x_0|V\rangle.$$

If $S_L = 1$, the only non-zero transition rate is $(h_L)_{1;1}$, therefore

$$
\begin{aligned}
&\sum_{\tau_L} (h_L)_{S_L;\,\tau_L} f_L(S_1,\ldots,S_{L-1},\tau_L) \\
&= \sum_{\tau_L} (h_L)_{1;\,\tau_L} f_L(S_1,\ldots,S_{L-1},\tau_L) = (h_L)_{1;1} f_L(S_1,\ldots,S_{L-1},1) \\
&= -\beta \langle W| \prod_{j=1}^{L-1} \Big(S_j D + \big(1-S_j\big) E \Big) D|V\rangle, \\
&\quad - x_{S_L} f_{L-1}(S_1,\ldots,S_{L-1}) = -x_1 \langle W| \prod_{j=1}^{L-1} \Big(S_j D + \big(1-S_j\big) E \Big) |V\rangle, \\
&\therefore \quad \beta D|V\rangle = x_1|V\rangle.
\end{aligned}
$$

Overall, one gets

$$\beta D|V\rangle = x_1|V\rangle = -x_0|V\rangle. \tag{2.1.11}$$

By choosing $x_1 = -x_0 = 1$, (2.1.9, 2.1.10, 2.1.11) is exactly (2.1.3). One may argue that in fact $x_1 = -x_0$ can be any non-zero constant C. It is true, however, one can always change the value of $x_1 = -x_0$ to be 1 by multiplying D and E by $\frac{1}{C}$. By now x_0, x_1 have been found, which implies that (2.1.6-2.1.8) hold, and $\dfrac{dP_L(S_1,S_2,\ldots,S_L)}{dt} = 0$. In other words, TASEP can be in steady state. Lemma 2.1 is proved. □

In the steady state, let $\langle S_i \rangle$ denote the average occupation of site i over time, $\langle S_i S_j \rangle$

represents the correlation between site i and site j, J represents the current which is the average number of particles passing through the lattice per unit time, then:

Lemma 2.2. ***The steady state formulation.***

$$\langle S_i \rangle = \frac{\langle W|C^{i-1}DC^{L-i}|V\rangle}{\langle W|C^L|V\rangle}, \tag{2.1.12}$$

$$\langle S_i S_j \rangle = \frac{\langle W|C^{i-1}DC^{j-i-1}DC^{L-j}|V\rangle}{\langle W|C^L|V\rangle}, \tag{2.1.13}$$

$$J = \frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle}, \tag{2.1.14}$$

where $C = D + E = DE$.

Proof. (see [44]) Based on the definition of average occupation, it can be shown that:

$$\begin{aligned}
\langle S_i \rangle &= 1 \cdot \sum_{S_1=0,1} \cdots \sum_{S_{j\neq i}=0,1} \cdots \sum_{S_L=0,1} P_L(S_1,\ldots,S_i=1,\ldots,S_L) \\
&= \left(\sum_{S_1=0,1} \cdots \sum_{S_{j\neq i}=0,1} \cdots \sum_{S_L=0,1} f_L(S_1,\ldots,S_i=1,\ldots,S_L) \right) \Big/ Z_L \\
&= \frac{\sum_{S_1=0,1}\cdots\sum_{S_{j\neq i}=0,1}\cdots\sum_{S_L=0,1} f_L(S_1,\ldots,S_i=1,\ldots,S_L)}{\sum_{S_1=0,1}\cdots\sum_{S_i=0,1}\cdots\sum_{S_L=0,1} f_L(S_1,\ldots,S_i,\ldots,S_L)} \\
&= \frac{\sum_{S_1=0,1}\cdots\sum_{S_{j\neq i}=0,1}\cdots\sum_{S_L=0,1}\langle W|\prod_{k=1}^{L}\left(S_k D+(1-S_k)E\right)|V\rangle}{\sum_{S_1=0,1}\cdots\sum_{S_i=0,1}\cdots\sum_{S_L=0,1}\langle W|\prod_{k=1}^{L}\left(S_k D+(1-S_k)E\right)|V\rangle} \\
&= \frac{\langle W|(D+E)^{i-1}D(D+E)^{L-i}|V\rangle}{\langle W|(D+E)^{L}|V\rangle} = \frac{\langle W|C^{i-1}DC^{L-i}|V\rangle}{\langle W|C^L|V\rangle}.
\end{aligned}$$

The correlation $\langle S_i S_j \rangle$, $i < j$ can be derived:

$$\begin{aligned}
&\langle S_i S_j \rangle \\
&= \sum_{S_1=0,1} \cdots \sum_{S_i=0,1} \cdots \sum_{S_j=0,1} \cdots \sum_{S_L=0,1} S_i S_j P_L\left(S_1,\ldots,S_i,\ldots,S_j,\ldots,S_L\right)
\end{aligned}$$

$$
\begin{aligned}
&= \left(\sum_{S_1=0,1} \cdots \sum_{S_i=1} \cdots \sum_{S_j=1} \cdots \sum_{S_L=0,1} f_L\left(S_1,\ldots,S_i=1,\ldots,S_j=1,\ldots,S_L\right) \right) \Big/ Z_L \\
&= \frac{\sum_{S_1=0,1}\cdots\sum_{S_i=1}\cdots\sum_{S_j=1}\cdots\sum_{S_L=0,1} f_L\left(S_1,\ldots,S_i=1,\ldots,S_j=1,\ldots,S_L\right)}{\sum_{S_1=0,1}\cdots\sum_{S_i=0,1}\cdots\sum_{S_j=0,1}\cdots\sum_{S_L=0,1} f_L\left(S_1,\ldots,S_i,\ldots,S_j,\ldots,S_L\right)} \\
&= \frac{\sum_{S_1=0,1}\cdots\sum_{S_i=1}\cdots\sum_{S_j=1}\cdots\sum_{S_L=0,1}\langle W|\prod_{k=1}^{L}\left(S_k D+(1-S_k)E\right)|V\rangle}{\sum_{S_1=0,1}\cdots\sum_{S_i=0,1}\cdots\sum_{S_j=0,1}\cdots\sum_{S_L=0,1}\langle W|\prod_{k=1}^{L}\left(S_k D+(1-S_k)E\right)|V\rangle} \\
&= \frac{\langle W|\left(D+E\right)^{i-1}D\left(D+E\right)^{j-i-1}D\left(D+E\right)^{L-j}|V\rangle}{\langle W|\left(D+E\right)^{L}|V\rangle} \\
&= \frac{\langle W|C^{i-1}DC^{j-i-1}DC^{L-j}|V\rangle}{\langle W|C^{L}|V\rangle}.
\end{aligned}
$$

In the steady state, the current J is defined as:

$$J = \langle \alpha(1-S_1)\rangle = \ldots = \langle S_i(1-S_{i+1})\rangle = \ldots = \langle \beta S_L\rangle.$$

Without loss of generality, current is derived as follows:

$$
\begin{aligned}
J &= \langle S_i(1-S_{i+1})\rangle = \langle S_i\rangle - \langle S_i S_{i+1}\rangle \\
&= \frac{\langle W|C^{i-1}DC^{L-i}|V\rangle}{\langle W|C^{L}|V\rangle} - \frac{\langle W|C^{i-1}DC^{i+1-i-1}DC^{L-i-1}|V\rangle}{\langle W|C^{L}|V\rangle} \\
&= \frac{\langle W|C^{i-1}DC^{L-i}|V\rangle - \langle W|C^{i-1}DDC^{L-i-1}|V\rangle}{\langle W|C^{L}|V\rangle} \\
&= \frac{\langle W|C^{i-1}D(C-D)C^{L-i-1}|V\rangle}{\langle W|C^{L}|V\rangle} = \frac{\langle W|C^{i-1}DEC^{L-i-1}|V\rangle}{\langle W|C^{L}|V\rangle} \\
&= \frac{\langle W|C^{i-1}CC^{L-i-1}|V\rangle}{\langle W|C^{L}|V\rangle} = \frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^{L}|V\rangle}.
\end{aligned}
$$

Lemma 2.2 is proved. □

Lemma 2.3. ***Expression of*** $\langle W|C^N|V\rangle$ ***and*** DC^n.

$$\langle W|C^N|V\rangle = \sum_{p=1}^{N} \frac{p(2N-1-p)!}{N!(N-p)!} \frac{\left(1/\beta\right)^{p+1} - \left(1/\alpha\right)^{p+1}}{\left(1/\beta\right) - \left(1/\alpha\right)} \tag{2.1.15}$$

$$DC^n = \sum_{p=0}^{n-1} \frac{(2p)!}{p!(p+1)!} C^{n-p} + \sum_{p=2}^{n+1} \frac{(p-1)(2n-p)!}{n!(n+1-p)!} D^p. \tag{2.1.16}$$

Proof. It can be shown that for $0 \leqslant p \leqslant n+1$,

$$\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!} - \frac{(p+1)(2n-p)!}{(n+1)!(n-p)!} = \frac{(p-1)(2n-p)!}{n!(n+1-p)!} \tag{2.1.17}$$

since

$$\begin{aligned}
&\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!} - \frac{(p+1)(2n-p)!}{(n+1)!(n-p)!} \\
=&\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!} - \frac{(p+1)(2n-p)!(n+1-p)}{(n+1)!(n+1-p)!} \\
=&\frac{\big(p(2n+1-p)-(p+1)(n+1-p)\big)(2n-p)!}{(n+1)!(n+1-p)!} \\
=&\frac{(n+1)(p-1)(2n-p)!}{(n+1)!(n+1-p)!} = \frac{(p-1)(2n-p)!}{n!(n+1-p)!},
\end{aligned}$$

and for $n \geqslant 0$, using the formula of the sum for a geometric series of matrix:

$$D^n C = C + \sum_{q=2}^{n+1} D^q = \frac{D^{n+2} - D^2}{D-I} + C = \frac{D^{n+2} - D}{D-I} + E, \tag{2.1.18}$$

note that $\dfrac{1}{D-I}$ means $(D-I)^{-1}$ here, namely, matrix inverse,

$$\begin{aligned}
\because \quad & D^n C = D^n(D+E) = D^{n+1} + D^n E = D^{n+1} + D^{n-1} DE \\
=& D^{n+1} + D^{n-1}(D+E) = D^{n+1} + D^{n-1} C, \\
\therefore \quad & D^n C = D^{n+1} + D^{n-1} C = D^{n+1} + D^n + D^{n-2} C = \ldots \\
=& D^{n+1} + D^n + D^{n-1} + \ldots + D^2 + D^0 C = C + \sum_{q=2}^{n+1} D^q, \\
\text{and} \quad & C + \sum_{q=2}^{n+1} D^q = C + \frac{D^2(I - D^n)}{I - D} = \frac{D^{n+2} - D^2}{D-I} + C,
\end{aligned}$$

and
$$\frac{D^{n+2}-D^2}{D-I}+C=\frac{D^{n+2}-D}{D-I}+\frac{D-D^2}{D-I}+C$$
$$=\frac{D^{n+2}-D}{D-I}+\frac{D(I-D)}{D-I}+C=\frac{D^{n+2}-D}{D-I}-D+C=\frac{D^{n+2}-D}{D-I}+E.$$

It can be proved by mathematical induction that

$$C^n=\sum_{p=0}^{n}\frac{p(2n-1-p)!}{n!(n-p)!}\sum_{q=0}^{p}E^qD^{p-q}\qquad n\geqslant 1. \tag{2.1.19}$$

For $n=1$

$$\begin{aligned}C&=\sum_{p=0}^{1}\frac{p(2-1-p)!}{1!(1-p)!}\sum_{q=0}^{p}E^qD^{p-q}\\&=\frac{0(2-1-0)!}{1!(1-0)!}\sum_{q=0}^{0}E^qD^{0-q}+\frac{1(2-1-1)!}{1!(1-1)!}\sum_{q=0}^{1}E^qD^{1-q}\\&=0+\sum_{q=0}^{1}E^qD^{1-q}=E^0D^{1-0}+E^1D^{1-1}=D+E,\end{aligned}$$

it is true. Assume (2.1.19) is true for n, then

$$\begin{aligned}C^{n+1}&=C^nC=\sum_{p=0}^{n}\frac{p(2n-1-p)!}{n!(n-p)!}\sum_{q=0}^{p}E^qD^{p-q}C\\&=\sum_{p=0}^{n}\frac{p(2n-1-p)!}{n!(n-p)!}\sum_{q=0}^{p}E^q\left(\frac{D^{p-q+2}-D}{D-I}+E\right)\\&=\sum_{p=0}^{n}\frac{p(2n-1-p)!}{n!(n-p)!}\sum_{q=0}^{p}\left(E^{q+1}+E^q\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\\&=\sum_{p=0}^{n}\left(\frac{(p+1)(2n-p)!}{(n+1)!(n-p)!}-\frac{(p+2)(2n-p-1)!}{(n+1)!(n-p-1)!}\right)\\&\quad\times\sum_{q=0}^{p}\left(E^{q+1}+E^q\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\\&=\sum_{p=0}^{n}\frac{(p+1)(2n-p)!}{(n+1)!(n-p)!}\sum_{q=0}^{p}\left(E^{q+1}+E^q\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\end{aligned}$$

$$
\begin{aligned}
&\quad -\sum_{p=0}^{n}\frac{(p+2)(2n-p-1)!}{(n+1)!(n-p-1)!}\sum_{q=0}^{p}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\\
&=\frac{(0+1)(2n-0)!}{(n+1)!(n-0)!}\left(E^{0+1}+E^{0}\left(\frac{D^{0-0+2}-D}{D-I}\right)\right)\\
&\quad +\sum_{p=1}^{n}\frac{(p+1)(2n-p)!}{(n+1)!(n-p)!}\sum_{q=0}^{p}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\\
&\quad -\sum_{p=0}^{n}\frac{(p+2)(2n-p-1)!}{(n+1)!(n-p-1)!}\sum_{q=0}^{p}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q+2}-D}{D-I}\right)\right)\\
&=\frac{(2n)!}{(n+1)!n!}(E+D)\\
&\quad +\sum_{p=2}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\sum_{q=0}^{p-1}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q+1}-D}{D-I}\right)\right)\\
&\quad -\sum_{p=2}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\sum_{q=0}^{p-2}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q}-D}{D-I}\right)\right)\\
&=\frac{(2n)!}{(n+1)!n!}(E+D)+\sum_{p=2}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\times\\
&\quad \left\{\sum_{q=0}^{p-1}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q+1}-D}{D-I}\right)\right)-\sum_{q=0}^{p-2}\left(E^{q+1}+E^{q}\left(\frac{D^{p-q}-D}{D-I}\right)\right)\right\}\\
&=\frac{(2n)!}{(n+1)!n!}(E+D)+\sum_{p=2}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\times\left\{\left(\sum_{q=0}^{p-1}E^{q+1}-\sum_{q=0}^{p-2}E^{q+1}\right)\right.\\
&\quad \left.+\sum_{q=0}^{p-1}\left(E^{q}\frac{D^{p-q+1}-D}{D-I}-E^{q}\frac{D^{p-q}-D}{D-I}\right)+E^{p-1}\frac{D^{p-p+1}-D}{D-I}\right\}\\
&=\frac{(2n)!}{(n+1)!n!}(E+D)+\sum_{p=2}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\left(E^{p}+\sum_{q=0}^{p-1}E^{q}D^{p-q}\right)\\
&=\sum_{p=0}^{n+1}\frac{p(2n+1-p)!}{(n+1)!(n+1-p)!}\sum_{q=0}^{p}E^{q}D^{p-q},
\end{aligned}
$$

so (2.1.19) is proven by mathematical induction. Note that (2.1.18) is used for red parts, and (2.1.17) is used for blue parts. The indices labelled green are reconstructed

so that the terms can be combined. Furthermore, from (2.1.3),

$$\langle W|E^n = \langle W|EE^{n-1} = \frac{1}{\alpha}\langle W|E^{n-1} = \ldots = \frac{1}{\alpha^n}\langle W| \tag{2.1.20}$$

$$D^n|V\rangle = D^{n-1}D|V\rangle = D^{n-1}\frac{1}{\beta}|V\rangle = \ldots = \frac{1}{\beta^n}|V\rangle. \tag{2.1.21}$$

Using the expression for C^N, (2.1.19), and also (2.1.20, 2.1.21),

$$\begin{aligned}
\langle W|C^N|V\rangle &= \langle W|\sum_{p=0}^{N}\frac{p(2N-1-p)!}{N!(N-p)!}\sum_{q=0}^{p}E^qD^{p-q}|V\rangle \\
&= \sum_{p=0}^{N}\frac{p(2N-1-p)!}{N!(N-p)!}\sum_{q=0}^{p}\langle W|E^qD^{p-q}|V\rangle \\
&= \sum_{p=0}^{N}\frac{p(2N-1-p)!}{N!(N-p)!}\sum_{q=0}^{p}\frac{1}{\alpha^q}\frac{1}{\beta^{p-q}}\langle W|V\rangle \\
&= \sum_{p=0}^{N}\frac{p(2N-1-p)!}{N!(N-p)!}\frac{\frac{1}{\beta^p}\left(1-\left(\frac{\beta}{\alpha}\right)^{p+1}\right)}{1-\frac{\beta}{\alpha}}\langle W|V\rangle \\
&= \sum_{p=1}^{N}\frac{p(2N-1-p)!}{N!(N-p)!}\frac{\left(1/\beta\right)^{p+1}-\left(1/\alpha\right)^{p+1}}{\left(1/\beta\right)-\left(1/\alpha\right)}\langle W|V\rangle \qquad N\geqslant 1,
\end{aligned}$$

formula (2.1.15) is therefore derived with the normalization $\langle W|V\rangle = 1$ (this normalization is used afterwards). Lemma 2.3 is proved. □

Based on Lemma 2.2 and Lemma 2.3, it is possible to obtain the exact expressions for the average occupation $\langle S_i\rangle$ and current J in the steady state even without knowing the forms for C, D, E and $\langle W|$, $|V\rangle$:

Lemma 2.4. ***The steady state density profile.***

$$\langle S_i\rangle_{i<L} = \sum_{p=0}^{l-1}\frac{(2p)!}{p!(p+1)!}\frac{\langle W|C^{L-p-1}|V\rangle}{\langle W|C^L|V\rangle} + \frac{\langle W|C^{i-1}|V\rangle}{\langle W|C^L|V\rangle}\sum_{p=2}^{l+1}\frac{(p-1)(2l-p)!}{l!(l+1-p)!}\beta^{-p},$$

$$\langle S_L\rangle = \frac{1}{\beta}\frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle}, \qquad J = \frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle},$$

where $l = L - i$ and $\langle W|C^N|V\rangle$ is given in Lemma 2.3.

Proof. The expression for current J is given by (2.1.14). Using the expression for average occupation $\langle S_i\rangle$ and (2.1.16), one can obtain:

for $l \geqslant 1$:

$$\begin{aligned}
\langle S_i\rangle &= \frac{\langle W|C^{i-1}DC^{L-i}|V\rangle}{\langle W|C^L|V\rangle} = \frac{\langle W|C^{i-1}DC^{l}|V\rangle}{\langle W|C^L|V\rangle} \\
&= \frac{\langle W|C^{i-1}\sum_{p=0}^{l-1}\frac{(2p)!}{p!(p+1)!}C^{l-p}|V\rangle}{\langle W|C^L|V\rangle} + \frac{\langle W|C^{i-1}\sum_{p=2}^{l+1}\frac{(p-1)(2l-p)!}{l!(l+1-p)!}D^{p}|V\rangle}{\langle W|C^L|V\rangle} \\
&= \sum_{p=0}^{l-1}\frac{(2p)!}{p!(p+1)!}\frac{\langle W|C^{i-1}C^{l-p}|V\rangle}{\langle W|C^L|V\rangle} + \sum_{p=2}^{l+1}\frac{(p-1)(2l-p)!}{l!(l+1-p)!}\frac{\langle W|C^{i-1}D^{p}|V\rangle}{\langle W|C^L|V\rangle} \\
&= \sum_{p=0}^{l-1}\frac{(2p)!}{p!(p+1)!}\frac{\langle W|C^{L-p-1}|V\rangle}{\langle W|C^L|V\rangle} + \frac{\langle W|C^{i-1}|V\rangle}{\langle W|C^L|V\rangle}\sum_{p=2}^{l+1}\frac{(p-1)(2l-p)!}{l!(l+1-p)!}\beta^{-p}
\end{aligned}$$

for $l = 0$ i.e., $i = L$:

$$\langle S_L\rangle = \frac{\langle W|C^{L-1}DC^{L-L}|V\rangle}{\langle W|C^L|V\rangle} = \frac{\langle W|C^{L-1}D|V\rangle}{\langle W|C^L|V\rangle} = \frac{1}{\beta}\frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle}.$$

Lemma 2.4 is proved. □

TASEP can be in a steady state in long time limit and the exact steady state solution is obtained in this section.

2.2 Mean Field Steady State Solution

One can see that the computational price to solve the exact steady state is massive, especially for large systems. For most applications of TASEP, systems are large. Therefore, one has to seek an approximation for the exact steady state solutions for large systems. In this section, we will follow Derrida *et al.* [43] to derive the mean field

solution for large systems. To derive the mean field solution, one starts from time dependent updating of site occupations and four phases will be defined.

Given the occupation of site i, $i \in [2, L-1]$ at time t is $S_i(t)$, then the occupation at time $t+1$ is given by:

$$S_i(t+1) = \begin{cases} S_i(t) + S_{i-1}(t)\left(1 - S_i(t)\right) & \text{if site } i-1 \text{ is chosen to update,} \\ \underbrace{S_i(t) - S_i(t)\left(1 - S_{i+1}(t)\right)}_{=S_i(t)S_{i+1}(t)} & \text{if site } i \text{ is chosen to update,} \\ S_i(t), & \text{otherwise.} \end{cases}$$

The particle reservoir, together with L sites are randomly chosen to update, so the probability of each of the first two events happening is $\frac{1}{L+1}$. Hence, the expected value of $S_i(t+1)$ is:

$$\begin{aligned} &\langle S_i(t+1)\rangle \\ =&\frac{\langle S_i(t)\rangle + \langle S_{i-1}(t)\rangle - \langle S_i(t)S_{i-1}(t)\rangle}{L+1} + \frac{\langle S_i(t)S_{i+1}(t)\rangle}{L+1} + \left(1 - \frac{2}{L+1}\right)\langle S_i(t)\rangle \\ =&\langle S_i(t)\rangle + \frac{\langle S_{i-1}(t)\rangle - \langle S_i(t)\rangle + \langle S_i(t)S_{i+1}(t)\rangle - \langle S_{i-1}(t)S_i(t)\rangle}{L+1}. \end{aligned}$$

Consider the occupation of site 1 at time $t+1$:

$$S_1(t+1) = \begin{cases} \underbrace{\begin{cases} 1 \text{ with associated prob. } \dfrac{S_1(t) + \alpha(1 - S_1(t))}{L+1}, \\ 0 \text{ with associated prob. } \dfrac{(1 - S_1(t))(1-\alpha)}{L+1}, \end{cases}}_{\text{reservoir is chosen}} \\ S_1(t)S_2(t) \quad \text{site 1 is chosen,} \\ S_1(t) \quad \text{otherwise.} \end{cases}$$

Both the particle resevoir and site 1 are chosen with associated probability $\frac{1}{L+1}$.

Hence, the expected value of $S_1(t+1)$ is:

$$\begin{aligned}
&\langle S_1(t+1)\rangle \\
=&\frac{\langle S_1(t)\rangle + \alpha(1-\langle S_1(t)\rangle)}{L+1} + \frac{\langle S_1(t)S_2(t)\rangle}{L+1} + \langle S_1(t)\rangle\left(1-\frac{2}{L+1}\right) \\
=&\langle S_1(t)\rangle + \frac{\alpha(1-\langle S_1(t)\rangle) + \langle S_1(t)S_2(t)\rangle - \langle S_1(t)\rangle}{L+1}.
\end{aligned}$$

Consider the occupation of site L at time t:

$$S_L(t+1) = \begin{cases} S_L(t) + S_{L-1}(t)\left(1-S_L(t)\right), \text{ site } L-1 \text{ is chosen} \\ \underbrace{\begin{cases} 1 \text{ with associated prob. } \dfrac{S_L(t)(1-\beta)}{L+1}, \\ 0 \text{ with prob. } \dfrac{1-S_L(t)+S_L(t)\beta}{L+1}, \end{cases}}_{\text{site } L \text{ is chosen}} \\ S_L(t), \text{ otherwise.} \end{cases}$$

The probabilities of site $L-1$ and L being chosen are both $\dfrac{1}{L+1}$. Hence, the expected value of $S_L(t+1)$ is:

$$\begin{aligned}
&\langle S_L(t+1)\rangle \\
=&\frac{\langle S_L(t)\rangle + \langle S_{L-1}(t)\rangle - \langle S_{L-1}(t)S_L(t)\rangle}{L+1} + \frac{\langle S_L(t)(1-\beta)\rangle}{L+1} + \langle S_L(t)\rangle\left(1-\frac{2}{L+1}\right) \\
=&\langle S_L(t)\rangle + \frac{\langle S_{L-1}(t)\rangle - \langle S_{L-1}(t)S_L(t)\rangle - \beta\langle S_L(t)\rangle}{L+1}.
\end{aligned}$$

To summarise,

$$\begin{aligned}
\langle S_1(t+1)\rangle - \langle S_1(t)\rangle &= \frac{\alpha(1-\langle S_1(t)\rangle) + \langle S_1(t)S_2(t)\rangle - \langle S_1(t)\rangle}{L+1}, \\
\langle S_i(t+1)\rangle - \langle S_i(t)\rangle &= \frac{\langle S_{i-1}(t)\rangle - \langle S_i(t)\rangle + \langle S_i(t)S_{i+1}(t)\rangle - \langle S_{i-1}(t)S_i(t)\rangle}{L+1}, \\
\langle S_L(t+1)\rangle - \langle S_L(t)\rangle &= \frac{\langle S_{L-1}(t)\rangle - \langle S_{L-1}(t)S_L(t)\rangle - \beta\langle S_L(t)\rangle}{L+1},
\end{aligned} \tag{2.2.1}$$

where $i = 2, \ldots, L-1$.

Mean field approximation is widely used to simplify large individual interacting systems, see [39, 137]. The essence of the mean field approximation is to neglect the correlation between the individuals, i.e., to set $\langle S_i S_{i+1} \rangle = \langle S_i \rangle \langle S_{i+1} \rangle$. Note that the mean field approximation agrees well to the exact solution for sufficient large L. It has been shown that TASEP can be in steady state. In the steady state, $\langle S_i \rangle$ remains unchanged. Thus the left hand sides of (2.2.1) are 0. Setting $\rho_i = \langle S_i \rangle$ and applying the mean field approximation, one gets:

$$\rho_i - \rho_i \rho_{i+1} = \rho_{i-1} - \rho_{i-1} \rho_i, \tag{2.2.2}$$

$$\alpha(1-\rho_1) = \rho_1 - \rho_1 \rho_2, \tag{2.2.3}$$

$$\beta \rho_L = \rho_{L-1} - \rho_{L-1} \rho_L. \tag{2.2.4}$$

Setting $\rho_i - \rho_i \rho_{i+1} = \rho_{i-1} - \rho_{i-1} \rho_i = J$ yields a recursion:

$$\rho_{i+1} = 1 - \frac{J}{\rho_i}, \tag{2.2.5}$$

where J is a constant to be determined later. In fact, J is the current under mean field approximation since:

$$J = \langle S_i \rangle - \langle S_i S_{i+1} \rangle = \rho_i - \rho_i \rho_{i+1}.$$

Recursion (2.2.5) has two fixed points given by the roots of the following function:

$$\rho = 1 - \frac{J}{\rho}, \quad \text{i.e.,} \quad y(\rho) = \rho^2 - \rho + J.$$

The fixed points $\rho_\pm = \frac{1}{2}\left(1 \pm \sqrt{1-4J}\right)$ exist provided $J \leqslant \frac{1}{4}$. The stability of the fixed point is determined by $\frac{d}{d\rho}\left(1 - \frac{J}{\rho}\right) = \frac{J}{\rho^2}$. If $\left|\frac{J}{\rho^2}\right| < 1$, then the fixed point is

stable, otherwise it is unstable. The fixed point ρ_- is unstable since:

$$\begin{aligned}\frac{J}{\rho_-^2}-1 &= \frac{J}{\left(\frac{1}{2}\left(1-\sqrt{1-4J}\right)\right)^2}-1 = \frac{4J-1+2\sqrt{1-4J}-\left(\sqrt{1-4J}\right)^2}{\left(1-\sqrt{1-4J}\right)^2} \\ &= \frac{\sqrt{1-4J}\left(2-2\sqrt{1-4J}\right)}{\left(1-\sqrt{1-4J}\right)^2} > 0, \qquad \text{i.e., } \frac{J}{\rho_-^2} > 1.\end{aligned}$$

The fixed point ρ_+ is stable since:

$$\begin{aligned}\frac{J}{\rho_+^2}-1 &= \frac{J}{\left(\frac{1}{2}\left(1+\sqrt{1-4J}\right)\right)^2}-1 = \frac{4J-1-2\sqrt{1-4J}-\left(\sqrt{1-4J}\right)^2}{\left(1+\sqrt{1-4J}\right)^2} \\ &= \frac{\sqrt{1-4J}\left(-2-2\sqrt{1-4J}\right)}{\left(1+\sqrt{1-4J}\right)^2} < 0, \qquad \text{i.e., } 0 < \frac{J}{\rho_+^2} < 1.\end{aligned}$$

Using the facts that $J = \dfrac{1-(\rho_+-\rho_-)^2}{4}$ and $\rho_+ + \rho_- = 1$, a new recursion can be defined. Let

$$\Psi_i := \frac{\rho_i - \rho_-}{\rho_i - \rho_+}, \tag{2.2.6}$$

then:

$$\begin{aligned}\Psi_{i+1} = \frac{\rho_{i+1}-\rho_-}{\rho_{i+1}-\rho_+} &= \frac{1-\frac{J}{\rho_i}-\rho_-}{1-\frac{J}{\rho_i}-\rho_+} = \frac{\rho_i - J - \rho_-\rho_i}{\rho_i - J - \rho_+\rho_i} = \frac{(1-\rho_-)\rho_i - J}{(1-\rho_+)\rho_i - J} \\ &= \frac{\rho_+\rho_i - \frac{1-(\rho_+-\rho_-)^2}{4}}{\rho_-\rho_i - \frac{1-(\rho_+-\rho_-)^2}{4}} \\ &= \frac{4\rho_+\rho_i - 1 + (\rho_+-\rho_-)^2}{4\rho_-\rho_i - 1 + (\rho_+-\rho_-)^2},\end{aligned}$$

from (2.2.6), one gets $\rho_i = \dfrac{\rho_- - \rho_+\Psi_i}{1-\Psi_i}$, then:

$$\begin{aligned}
\Psi_{i+1} &= \frac{4\rho_+\dfrac{\rho_- - \rho_+\Psi_i}{1-\Psi_i} - 1 + (\rho_+ - \rho_-)^2}{4\rho_-\dfrac{\rho_- - \rho_+\Psi_i}{1-\Psi_i} - 1 + (\rho_+ - \rho_-)^2} \\
&= \frac{4\rho_+(\rho_- - \rho_+\Psi_i) - 1 + \Psi_i + (\rho_+ - \rho_-)^2 - (\rho_+ - \rho_-)^2\Psi_i}{4\rho_-(\rho_- - \rho_+\Psi_i) - 1 + \Psi_i + (\rho_+ - \rho_-)^2 - (\rho_+ - \rho_-)^2\Psi_i} \\
&= \frac{4\rho_+\rho_- + (\rho_+ - \rho_-)^2 - 1 - 4\rho_+^2\Psi_i + \Psi_i - (\rho_+ - \rho_-)^2\Psi_i}{4\rho_-^2 - 1 + (\rho_+ - \rho_-)^2 - 4\rho_-\rho_+\Psi_i + \Psi_i - (\rho_+ - \rho_-)^2\Psi_i} \\
&= \frac{(\rho_+ + \rho_-)^2 - 1 + \Psi_i\left((1-2\rho_+)(1+2\rho_+) - (2\rho_+ - 1)^2\right)}{(2\rho_- - 1)(2\rho_- + 1) + (1-2\rho_-)^2 + \Psi_i\left(1 - (\rho_+ + \rho_-)^2\right)} \\
&= \frac{4\rho_+(1-2\rho_+)}{4\rho_-(2\rho_- - 1)}\Psi_i = \frac{\rho_+(2\rho_- - 1)}{\rho_-(2\rho_- - 1)}\Psi_i \\
&= \frac{\rho_+}{\rho_-}\Psi_i.
\end{aligned}$$

Therefore:

$$\Psi_i = \Psi_1\left(\frac{\rho_+}{\rho_-}\right)^{i-1} = \frac{(\rho_1 - \rho_-)\rho_+^{i-1}}{(\rho_1 - \rho_+)\rho_-^{i-1}},$$

and:

$$\begin{aligned}
\rho_i = \frac{\rho_- - \rho_+\Psi_i}{1-\Psi_i} &= \frac{\rho_- - \rho_+\dfrac{(\rho_1 - \rho_-)\rho_+^{i-1}}{(\rho_1 - \rho_+)\rho_-^{i-1}}}{1 - \dfrac{(\rho_1 - \rho_-)\rho_+^{i-1}}{(\rho_1 - \rho_+)\rho_-^{i-1}}} = \frac{\rho_-^i(\rho_1 - \rho_+) - \rho_+^i(\rho_1 - \rho_-)}{\rho_-^{i-1}(\rho_1 - \rho_+) - \rho_+^{i-1}(\rho_1 - \rho_-)} \\
&= \frac{-\rho_-^i\rho_+ + \rho_+^i\rho_- + \left(\rho_-^i - \rho_+^i\right)\rho_1}{-\rho_-^{i-1}\rho_+ + \rho_+^{i-1}\rho_- + \left(\rho_-^{i-1} - \rho_+^{i-1}\right)\rho_1} \\
&= \frac{\rho_+\rho_-\left(\rho_+^{i-1} - \rho_-^{i-1}\right) + \left(\rho_-^i - \rho_+^i\right)\rho_1}{\rho_+\rho_-\left(\rho_+^{i-2} - \rho_-^{i-2}\right) + \left(\rho_-^{i-1} - \rho_+^{i-1}\right)\rho_1}, \\
\text{i.e.,}\quad \rho_i &= \frac{-\rho_+\rho_-\left(\rho_+^{i-1} - \rho_-^{i-1}\right) + \left(\rho_+^i - \rho_-^i\right)\rho_1}{-\rho_+\rho_-\left(\rho_+^{i-2} - \rho_-^{i-2}\right) + \left(\rho_+^{i-1} - \rho_-^{i-1}\right)\rho_1},
\end{aligned} \tag{2.2.7}$$

where $\rho_\pm$ are fixed points. Replacing $\rho_1 - \rho_1\rho_2$ with J in (2.2.3) and $\rho_{L-1} - \rho_{L-1}\rho_L$

with J in (2.2.4), it yields:

$$\rho_1 = 1 - \frac{J}{\alpha}, \qquad \rho_L = \frac{J}{\beta}. \tag{2.2.8}$$

The mean field density profile is therefore given by (2.2.7, 2.2.8). Comparison between the mean field approximation and the exact solution is shown in Appendix A

It has been shown that ρ_- is unstable and ρ_+ is stable. After L iterations, recursion (2.2.5) has four typical cobwebs as shown in Figure 2.1. Note that in general, we consider L to be large but finite.

Case A: ρ_1 is close to the unstable fixed point ρ_-, i.e., $\rho_1 = \rho_- \pm \varepsilon(L)$, ρ_i deviates from ρ_- and $\rho_L < \rho_+$. $\varepsilon(L)$ is an infinitesimal.
Case B: $\rho_1 > \rho_-$, ρ_i approaches to the stable fixed points ρ_+, $\rho_L = \rho_+ \pm \varepsilon(L)$.
Case C: ρ_i deviates from ρ_- and approaches to ρ_+, $\rho_1 = \rho_- + \varepsilon(L)$, $\rho_L = \rho_+ - \varepsilon(L)$.
Case D: $J > \frac{1}{4}$ so there are no fixed points, $\rho_1 > \frac{1}{2}$ and $\rho_L < \frac{1}{2}$, most ρ_is stay around $\frac{1}{2}$. In fact, J has to be close to $\frac{1}{4}$, i.e., $J = \frac{1}{4} + \varepsilon(L)$. This is required by the precondition for mean field approximation: sufficient large systems. Otherwise, if $J \gg \frac{1}{4}$, cobweb Figure 2.1D gives small systems.

Given these four cases and the mean field expressions for ρ_1 and ρ_L, one is able to derive four phases depending on different values of α and β.

Lemma 2.5. *For case A,* $\alpha < \beta$, $\alpha < \frac{1}{2}$ *and* $\rho_1 \simeq \alpha$, $\rho_L \simeq \frac{\alpha(1-\alpha)}{\beta}$, $J \simeq \alpha(1-\alpha)$, $\rho \simeq \alpha$, *where* $\rho = \frac{\sum \rho_i}{L}$.

Proof. Now we consider Case A, ρ_1 is close to ρ_-:

$$\rho_1 = 1 - \frac{J}{\alpha} = \rho_- \pm \varepsilon(L) = \frac{1}{2}\left(1 - \sqrt{1-4J}\right) \pm \varepsilon(L)$$

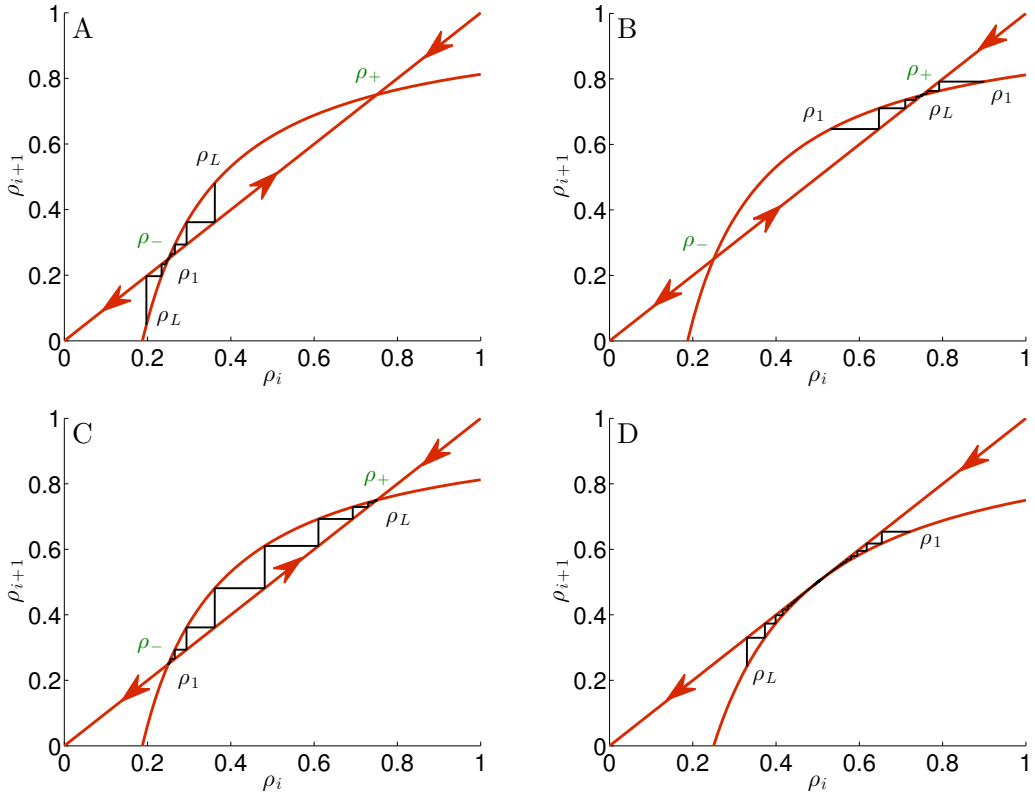


Figure 2.1: **Cobweb diagram of recursion 2.2.5.** Four typical cases of cobwebs with large but finite iteration are shown here. The red curve is $\rho_{i+1} = 1 - J/\rho_i$, and the red oblique line is $\rho_{i+1} = \rho_i$. $J = \dfrac{3}{16}$ in (A, B, C) and $J = \dfrac{1}{4} + 10^{-5}$ in (D).

yields $2J - \alpha \simeq \alpha\sqrt{1-4J} > 0$. Squaring both sides yields $J \simeq \alpha(1-\alpha)$, implies that $\rho_1 \simeq \alpha$ and $\rho_L \simeq \dfrac{\alpha(1-\alpha)}{\beta}$. Substituting $J \simeq \alpha(1-\alpha)$ into $2J - \alpha > 0$ yields:

$$2J - \alpha \simeq 2\alpha(1-\alpha) - \alpha > 0 \Longrightarrow \alpha < \frac{1}{2},$$

and from $\rho_L < \rho_+$:

$$\frac{\alpha(1-\alpha)}{\beta} < \frac{1}{2}\left(1 + \sqrt{1-4\alpha(1-\alpha)}\right) \Longrightarrow \alpha < \beta.$$

From Figure 2.1A, one can see that most ρ_i stay close to ρ_-, so $\rho \simeq \rho_-$, namely, $\rho \simeq \alpha$. The resulting mean field density profile for this phase is given in Figure 2.2A. □

Lemma 2.6. *For case B,* $\beta < \alpha, \beta < \frac{1}{2}$ *and* $\rho_1 \simeq 1 - \frac{\beta(1-\beta)}{\alpha}$, $\rho_L \simeq 1-\beta$, $J \simeq \beta(1-\beta)$, $\rho \simeq 1-\beta$.

Proof. Now we consider Case B, ρ_L is close to ρ_+:

$$\rho_L = \frac{J}{\beta} = \rho_+ \pm \varepsilon(L) = \frac{1}{2}\left(1 + \sqrt{1-4J}\right) \pm \varepsilon(L)$$

yields $2J - \beta \simeq \beta\sqrt{1-4J} > 0$. Squaring both sides gives $J \simeq \beta(1-\beta)$, implies that $\rho_1 \simeq 1 - \frac{\beta(1-\beta)}{\alpha}$ and $\rho_L = 1 - \frac{J}{\beta} \simeq 1-\beta$. Substituting $J \simeq \beta(1-\beta)$ into $2J-\beta > 0$ yields:

$$2J - \beta \simeq 2\beta(1-\beta) - \beta > 0 \Longrightarrow \beta < \frac{1}{2},$$

and from $\rho_1 > \rho_-$:

$$1 - \frac{\beta(1-\beta)}{\alpha} > \frac{1}{2}\left(1 - \sqrt{1-4\beta(1-\beta)}\right) \Longrightarrow \beta < \alpha.$$

From Figure 2.1B, one can see that most ρ_i stay close to ρ_+, so $\rho \simeq \rho_+$, namely, $\rho \simeq 1-\beta$. The resulting mean field density profile for this phase is given in Figure 2.2B. □

Lemma 2.7. *For case C,* $\alpha = \beta < \frac{1}{2}$ *and* $\rho_1 \simeq \alpha$, $\rho_L \simeq 1-\alpha$, $J \simeq \alpha(1-\alpha)$, $\rho \simeq \alpha$.

Proof. Now we consider Case C, ρ_1 is close to ρ_- and ρ_L is close to ρ_+:

$$\rho_1 = 1 - \frac{J}{\alpha} = \rho_- + \varepsilon(L) = \frac{1}{2}\left(1 - \sqrt{1-4J}\right) + \varepsilon(L),$$
$$\rho_L = \frac{J}{\beta} = \rho_+ - \varepsilon(L) = \frac{1}{2}\left(1 + \sqrt{1-4J}\right) - \varepsilon(L)$$

yields

$$\frac{2J}{\alpha} - 1 + 2\varepsilon(L) = \frac{2J}{\beta} - 1 + 2\varepsilon(L) = \sqrt{1-4J} > 0.$$

$\frac{2J}{\alpha} - 1 + 2\varepsilon(L) = \frac{2J}{\beta} - 1 + 2\varepsilon(L)$ has two solutions, $J = 0$ and $\alpha = \beta$. The trivial solution $J = 0$ is not physical relevant, so it is abandoned. $\frac{2J}{\alpha} - 1 + 2\varepsilon(L) = \sqrt{1 - 4J}$ yields $J \simeq \alpha(1-\alpha)$, implies that $\rho_1 \simeq \alpha$ and $\rho_L = \frac{\alpha(1-\alpha)}{\beta}$. Substituting $J \simeq \alpha(1-\alpha)$ into $\frac{2J}{\alpha} - 1 + 2\varepsilon(L) > 0$ gives $\alpha < \frac{1}{2}$.

The resulting mean field density profile for this phase is given in Figure 2.2C. The typical density profile has a sharp transition from a low density level, α, to a high density level, $1-\alpha$. The Case C thus is a special case that the Case A and Case B are coexistent. Furthermore, as we can see from Figure 2.1C and Figure 2.2C, due to the symmetry of the typical density profile, $\rho = \frac{1}{2}$. □

Lemma 2.8. *For case D,* $\alpha, \beta \geqslant \frac{1}{2}$ *and* $\rho_1 \simeq 1 - \frac{1}{4\alpha}$, $\rho_L \simeq \frac{1}{4\beta}$, $J \simeq \frac{1}{4}$, $\rho \simeq \frac{1}{2}$.

Proof. $J = \frac{1}{4} + \varepsilon(L)$ directly gives $J \simeq \frac{1}{4}$. Using (2.2.8), one is able to obtain $\rho_1 \simeq 1 - \frac{1}{4\alpha}$ and $\rho_L \simeq \frac{1}{4\beta}$. Moreover, from $\rho_1 > \frac{1}{2} > \rho_L$, one can easily obtain that $\alpha, \beta \geqslant \frac{1}{2}$. From Figure 2.1D, most ρ_i stay around $\frac{1}{2}$, so $\rho \simeq \frac{1}{2}$. The resulting mean field density profile for this phase is given in Figure 2.2D. □

The phases defined by Lemma 2.5-2.8 are shown in Figure 2.3. Subfigure (A) is for the average density. Subfigure (B) is for the current. The phase bounded by $\alpha < \beta$, $\alpha < \frac{1}{2}$ is called *low density phase* since the average density in this phase is relatively low. The phase bounded by $\beta < \alpha$, $\beta < \frac{1}{2}$ is called *high density phase* since the average density in this phase is relatively high. The line $\alpha = \beta < \frac{1}{2}$ is the boundary between the low density phase and the high density phase. It is referred to *coexistence line* or *shock phase* since it shares the properties of the low density phase and the high density phase. The average density is discontinuous when crosses over the coexistence line. In the phase bounded by $\alpha, \beta \geqslant \frac{1}{2}$, current reaches the maximal value, thus, this phase is called *maximal current phase*.

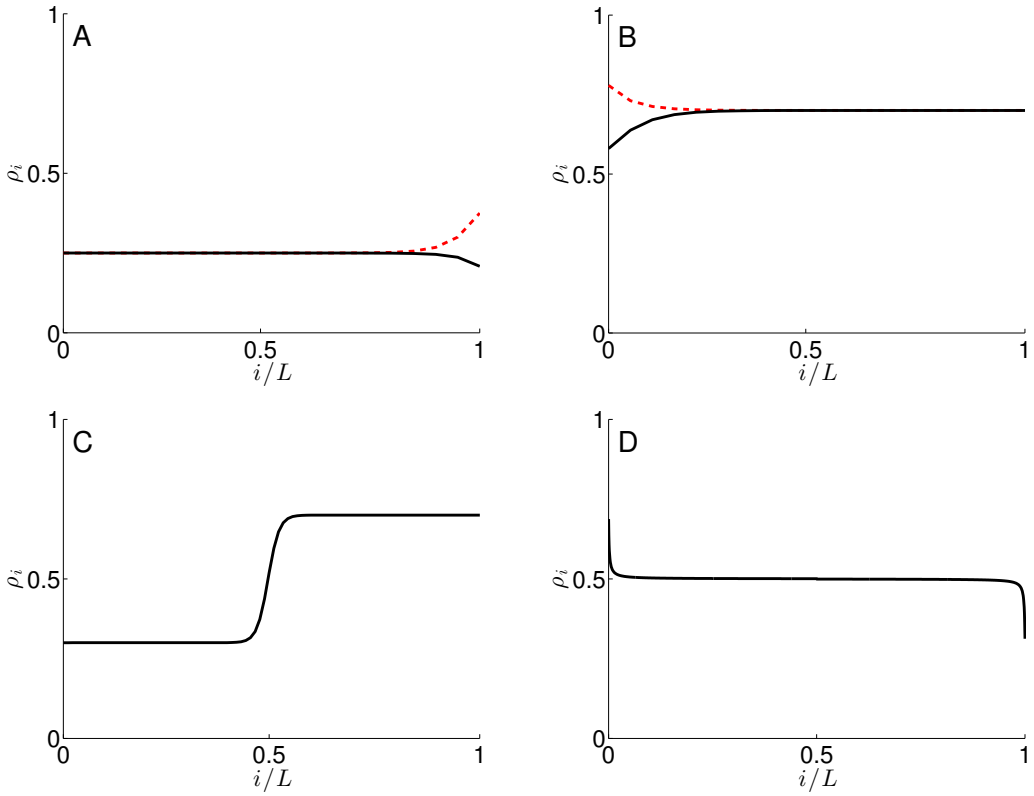
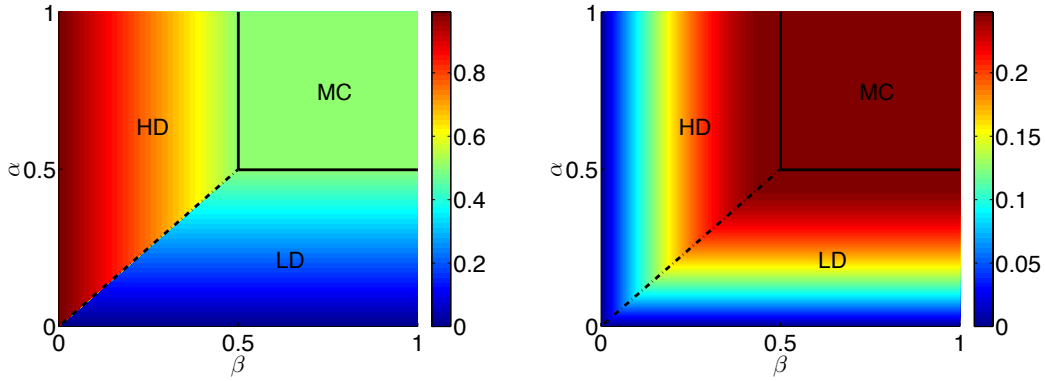


Figure 2.2: **Density profile of TASEP.** This figure shows typical density profiles for four phases. (A) Low density phase: $\alpha = 0.25, \beta = 0.5(\text{black})/\beta = 0.9(\text{red})$. (B) High density phase: $\alpha = 0.5(\text{black})/\alpha = 0.95(\text{red}), \beta = 0.3$. (C) Coexistence line: $\alpha = \beta = 0.3$. (D) Maximal current phase: $\alpha = \beta = 0.8$. The profile densities are given by (2.2.7, 2.2.8).



(a) ρ (b) J

Figure 2.3: **Illustration of the mean field solution.** (A): average density ρ. (B): current J. Four phases are displayed, the coexistence line (shock phase) is represented by the dashed line.

2.3 Transient Behaviour Approximation

We summarised the widely used steady state solution of TASEP in previous sections. However, transient behaviour of TASEP has attracted little attention and the corresponding theory is not well developed. Nagar *et al.* [108] gave a way to approximate the transient behaviour. The transient behaviour is important for some applications of TASEP, for example, modelling protein production which we will develop later. In this section, we will develop an approximation for the transient behaviour of TASEP based on Nagar *et al.* [108]. We will study a rate that particles enter the lattice denoted by J_1 and a rate that particles exit the lattice denoted by J_2. Let N_1 be the total number of particles entering the lattice during $[0,t]$ and N_2 be the total number of particles exiting the lattice during $[0,t]$, N_1 and N_2 are given by: $N_1 = \int_0^t J_1 dt$, $N_2 = \int_0^t J_2 dt$.

2.3.1 Low Density Phase

TASEP is in *low density phase* if $\alpha < \beta, \alpha < \frac{1}{2}$. A typical time evolution of the density profile of the low density phase is shown in Figure 2.4. The very first particle that binds to site 1 will take an expected time L to reach site L. Thus, the transient state of the TASEP can be divided into two stages. The first stage is during the time interval $[0,L]$, see the trajectories at $t = 300, 600$ for $L = 1000$ in Figure 2.4. The second stage is during the time interval $(L, t_c^{LD}]$, where t_c^{LD} is the time that the whole lattice accomplishes a steady state, namely, the global steady state (see the trajectory at $t = 1600$). In the first stage, no particle leaves the lattice, so $J_2 = 0$. A wave that travels from left to right is observed in Figure 2.4. The left part of the lattice reaches a steady state before the other part. In fact, the leftmost part of the lattice reaches the steady state very quickly. One can understand it by treating the leftmost part of the lattice as a sub system. This left sub system has a short lattice, a particle entry rate $\alpha < \frac{1}{2}$, and

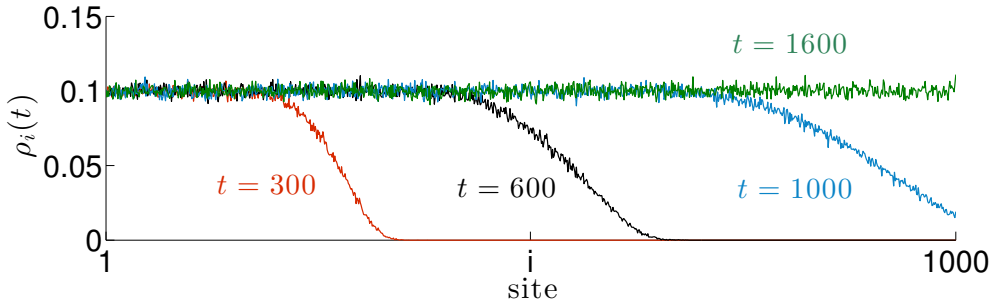


Figure 2.4: **Time evolution of density profile for low density phase.** Density profile $\rho_i(t)$ is plotted, different colours show different times, $t = 300/600/1000/1600$. Each of $\rho_i(t)$ is the average of 2000 simulations. We use $\alpha = 0.1, \beta = 0.8, L = 1000$. We will introduce simulation methods later.

a maximal exit rate, 1. Particle motion in the sub system is essentially not restricted by β, it is affected only by the 'hard core' repulsion between adjacent particles. Thus, the left sub system is an approximated TASEP in the low density phase and reaches a steady state quickly. J_1 in the first stage is determined by the current of the left sub system, which can be approximated by $\alpha(1-\alpha)$. In the second stage, J_1 is also $\alpha(1-\alpha)$ since the left sub system has been in the steady state. Particles start leaving the lattice, thus J_2 increases from 0 until the steady state $\alpha(1-\alpha)$ is reached. As a first approximation, we set $J_2 = \dfrac{\alpha(1-\alpha)}{2}$ during the second stage. After t_c^{LD}, the whole lattice is in the global steady state of the low density phase, thus $J_1 = J_2 = \alpha(1-\alpha)$. Unlike us, Nagar *et al.* [108] ignored the transition of J_2 from 0 to $\alpha(1-\alpha)$. They assumed that J_2 is 0 before the global steady state is reached, and immediately becomes $\alpha(1-\alpha)$ after that.

At time t_c^{LD}, we suppose that the lattice is in a global steady state, i.e., the net number of particles added to the lattice in $[0, t_c^{LD}]$ can be approximated by the total number of particles on the lattice in the steady state:

$$\alpha(1-\alpha)L + \left(\alpha(1-\alpha) - \frac{\alpha(1-\alpha)}{2}\right)\left(t_c^{LD} - L\right) = \alpha L \quad \Longrightarrow \quad t_c^{LD} = \frac{1+\alpha}{1-\alpha}L.$$

The red term is the net number of particles added to the lattice during the first stage,

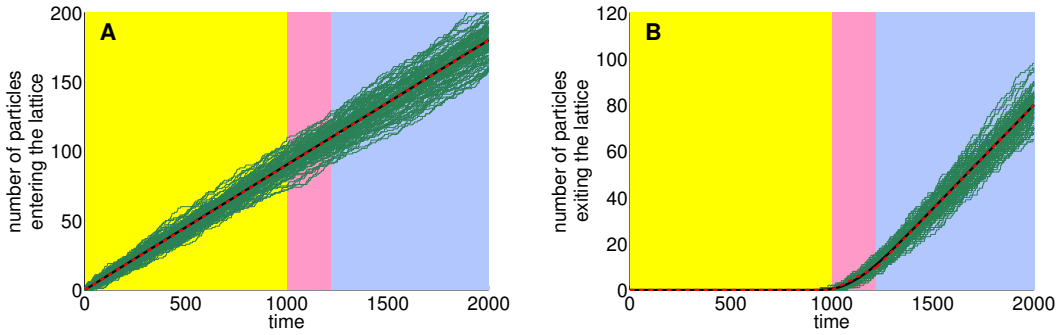


Figure 2.5: **Transient state approximation for low density phase.** (A) Number of particles entering the lattice and (B) number of particles exiting lattice are plotted as functions of time. Red dashed lines are approximations (2.3.2) and black lines are stochastic averages from 2000 simulations. 20 realizations of simulation are plotted (green). Yellow region is the first stage and pink region is the second stage. In purple region, TASEP is in the steady state. We use $\alpha = 0.1$, $\beta = 0.8$, $L = 1000$.

and the blue term is the net number of particles added to the lattice during the second stage. The green term, αL, is the number of particles on the lattice in the steady state, which is given by the product of average density in the steady state of the low density phase and the length of the lattice (see Lemma 2.5). In summary, in the low density phase:

$$J_1 = \alpha(1-\alpha), \quad t > 0; \qquad J_2 = \begin{cases} 0, & 0 < t < L, \\ \dfrac{\alpha(1-\alpha)}{2}, & L \leqslant t < t_c^{LD}, \\ \alpha(1-\alpha), & t \geqslant t_c^{LD}. \end{cases} \tag{2.3.1}$$

Following the approximation (2.3.1) above, we have the total number of particles entering the lattice in $[0,t]$ and the total number of particles exiting the lattice in $[0,t]$:

$$N_1 = \alpha(1-\alpha)t; \quad N_2 = \begin{cases} 0, & 0 < t < L, \\ \dfrac{\alpha(1-\alpha)}{2}(t-L), & L \leqslant t < t_c^{LD}, \\ \alpha(1-\alpha)t - \alpha L, & t \geqslant t_c^{LD}. \end{cases} \tag{2.3.2}$$

The transient state approximation for the low density phase is then given by (2.3.1), (2.3.2), and it is shown in Figure 2.5. The approximation (2.3.2) agrees well with the average of simulations.

2.3.2 High Density Phase

TASEP is in *high density phase* if $\beta < \alpha, \beta < \frac{1}{2}$. Typical time evolution of density profile in the high density phase is shown in Figure 2.6. In this case, we find that the right part of the lattice reaches its steady state before the left part, and a wave travels from right to left until the whole lattice reaches the global steady state. This is because the exit rate β is the limiting rate, and hence particles queue on the lattice. What is interesting is the density profile evolution of the left end of the lattice. Particles enter an empty lattice at $t = 0$. The leftmost part of the lattice can be regarded as a small sub system with an entry rate α and a maximal exit rate, 1. Based on the basic theory, the sub system will be in the low density phase if $\alpha < \frac{1}{2}$, and will be in the maximal current phase if $\alpha \geqslant \frac{1}{2}$. These two cases are illustrated in Figure 2.6. As we can see, the density profile of the leftmost part of the lattice for $\alpha < \frac{1}{2}$ is flat while the density profile of the leftmost part of the lattice for $\alpha \geqslant \frac{1}{2}$ is not. We now consider $\alpha < \frac{1}{2}$ and $\alpha \geqslant \frac{1}{2}$ separately.

2.3.2.1 $\alpha < 1/2$

We divide the transient state of the TASEP in the high density phase with $\alpha < \frac{1}{2}$ into two stages, $[0, L]$ and $(L, t_{c1}^{HD}]$, where t_{c1}^{HD} is the time that the whole lattice reaches the global steady state. Motivated by the simulations shown in Figure 2.6, we consider the leftmost part of the lattice to be a sub system in the low density phase during $[0, t_{c1}^{HD}]$ by neglecting the rapid transition to the high density phase at the very end of $[0, t_{c1}^{HD}]$.

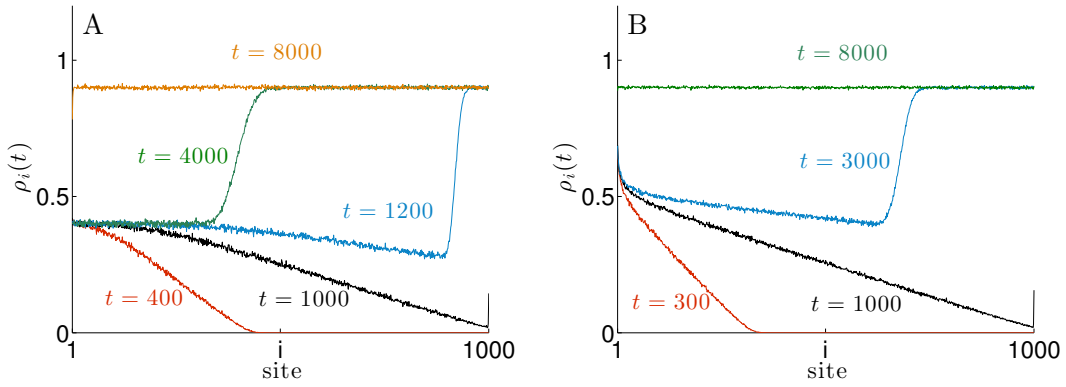


Figure 2.6: **Time evolution of density profile for high density phase.** Density profile $\rho_i(t)$ is plotted, different colours show different times. Each of $\rho_i(t)$ is the average of 2000 simulations. (A) $\alpha = 0.4$, $\beta = 0.1$, $L = 1000$ and (B) $\alpha = 0.8$, $\beta = 0.1$, $L = 1000$.

Therefore, $J_1 = \alpha(1-\alpha)$ in $[0, t_{c1}^{HD}]$, and $J_1 = \beta(1-\beta)$ after t_{c1}^{HD}. In the first stage $[0, L]$, no particle leaves the lattice, so $J_2 = 0$. Neglecting the rapid transition of J_2 from 0 to the steady state, J_2 is set to be $\beta(1-\beta)$ after $t = L$.

Using the same argument we used to define t_c^{LD}, we are able to find t_{c1}^{HD}:

$$\begin{aligned} & \alpha(1-\alpha)L + \left(\alpha(1-\alpha) - \beta(1-\beta)\right)\left(t_{c1}^{HD} - L\right) = (1-\beta)L \\ \Longrightarrow \quad & t_{c1}^{HD} = \frac{(1-\beta)^2 L}{\alpha(1-\alpha) - \beta(1-\beta)}. \end{aligned}$$

The red term is the net number of particles added to the lattice during the first stage, and the blue term is the net number of particles added to the lattice during the second stage. The green term, $(1-\beta)L$, is the number of particles on the lattice in steady state, which is given by the product of average density in the steady state and the length of the lattice. In summary, in the high density phase $\left(\text{for } \alpha < \frac{1}{2}\right)$:

$$J_1 = \begin{cases} \alpha(1-\alpha), & t < t_{c1}^{HD}, \\ \beta(1-\beta), & t \geqslant t_{c1}^{HD}. \end{cases} \qquad J_2 = \begin{cases} 0, & t < L, \\ \beta(1-\beta), & t \geqslant L. \end{cases} \tag{2.3.3}$$

The total numbers of particles entering and exiting the lattice in $[0, t]$ are given by:

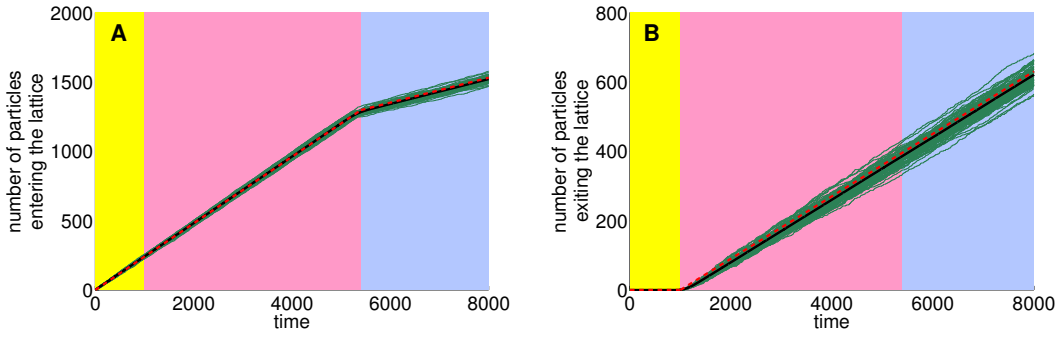


Figure 2.7: **Transient state approximation for high density phase: $\alpha < 1/2$.** (A) Number of particles entering the lattice and (B) number of particles exiting lattice are plotted as functions of time. Red dashed lines are approximations (2.3.4) and black lines are stochastic averages from 2000 simulations. 20 realizations of simulation are plotted (green). Yellow region is the first stage and pink region is the second stage. In purple region, TASEP is in the steady state. We use $\alpha = 0.4$, $\beta = 0.1$, $L = 1000$.

$$N_1 = \begin{cases} \alpha(1-\alpha)t, & t < t_{c1}^{HD}, \\ \beta(1-\beta)t + (1-\beta)^2 L, & t \geqslant t_{c1}^{HD}. \end{cases} \qquad N_2 = \begin{cases} 0, & t < L, \\ \beta(1-\beta)(t-L), & t \geqslant L. \end{cases} \tag{2.3.4}$$

The transient state approximation for the high density phase $\left(\text{for } \alpha < \frac{1}{2}\right)$ is then given by (2.3.3), (2.3.4), and is shown in Figure 2.7. As is clear from the Figure, the approximation (2.3.4) agrees well with the average of simulations.

2.3.2.2 $\alpha \geqslant 1/2$

Using the same argument we used for the case $\alpha < \frac{1}{2}$, we divide the transient state into two stages, $[0, L]$ and $(L, t_{c2}^{HD}]$, where t_{c2}^{HD} is the time that the lattice reaches the global steady state. The leftmost part of the lattice is treated as a sub system in the maximal current phase during $[0, t_{c2}^{HD}]$. Therefore, J_1 is approximated by $\frac{1}{4}$ in $[0, t_{c2}^{HD}]$, and $J_1 = \beta(1-\beta)$ after t_{c2}^{HD}. In the first stage, no particle leaves the lattice, so $J_2 = 0$. Neglecting the rapid transition of J_2 from 0 to the steady state, J_2 is set to be $\beta(1-\beta)$ after $t = L$. The net number of particles added to the lattice during $[0, t_{c2}^{HD}]$ is approximated by the total number of particles on the lattice in the steady state, we are able to find t_{c2}^{HD} by

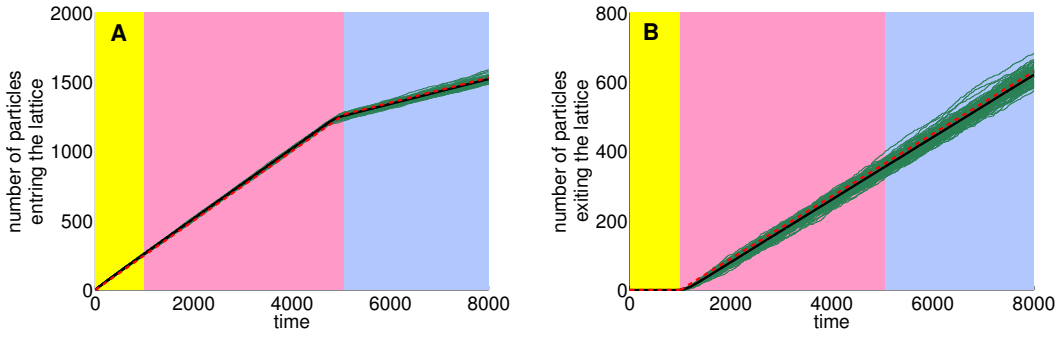


Figure 2.8: **Transient state approximation for high density phase:** $\alpha \geqslant 1/2$**.** (A) Number of particles entering the lattice and (B) number of particles exiting lattice are plotted as functions of time. Red dashed lines are approximations (2.3.6) and black lines are stochastic averages from 2000 simulations. 20 realizations of simulation are plotted (green).Yellow region is the first stage and pink region is the second stage. In purple region, TASEP is in the steady state. We use $\alpha = 0.8, \beta = 0.1, L = 1000$.

solving following equality:

$$\frac{1}{4}L + \left(\frac{1}{4} - \beta(1-\beta)\right)\left(t_{c2}^{HD} - L\right) = (1-\beta)L \quad \Longrightarrow \quad t_{c2}^{HD} = \frac{(3-4\beta)L}{1-4\beta(1-\beta)}.$$

The red term is the net number of particles added to the lattice during the first stage $[0, L]$, and the blue term is the net number of particles added to the lattice during the second stage $(L, t_{c2}^{HD}]$. The green term, $(1-\beta)L$, is the number of particles on the lattice in the steady state, which is given by the product of average density in the steady state and the length of the lattice. In summary, in the high density phase $\left(\text{for } \alpha \geqslant \frac{1}{2}\right)$:

$$J_1 = \begin{cases} \frac{1}{4}, & t < t_{c2}^{HD}, \\ \beta(1-\beta), & t \geqslant t_{c2}^{HD}. \end{cases} \qquad J_2 = \begin{cases} 0, & t < L, \\ \beta(1-\beta), & t \geqslant L. \end{cases} \tag{2.3.5}$$

The total numbers of particles entering and exiting the lattice in $[0, t]$ are:

$$N_1 = \begin{cases} \frac{1}{4}t, & t < t_{c2}^{HD}, \\ \beta(1-\beta)t + \frac{3-4\beta}{4}L, & t \geqslant t_{c2}^{HD}. \end{cases} \qquad N_2 = \begin{cases} 0, & t < L, \\ \beta(1-\beta)(t-L), & t \geqslant L. \end{cases} \tag{2.3.6}$$

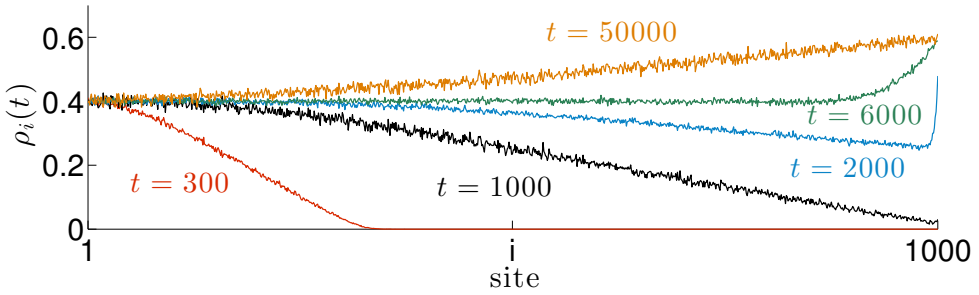


Figure 2.9: **Time evolution of density profile for coexistence line.** Density profile $\rho_i(t)$ is plotted, different colours show different times. Each of $\rho_i(t)$ is the average of 2000 simulations. We use $\alpha = 0.4$, $\beta = 0.4$, $L = 1000$.

The transient behaviour approximation for the high density phase $\left(\text{for } \alpha \geqslant \frac{1}{2}\right)$ is then given by (2.3.5), (2.3.6), and is shown in Figure 2.8. As we can see from the figure, the approximation (2.3.6) agrees well with the average of simulations.

2.3.3 Coexistence Line

TASEP is in *coexistence line* if $\alpha = \beta < \frac{1}{2}$. A typical time evolution of density profile of the coexistence line is shown in Figure 2.9. We find that the density profiles at different times intersect at the leftmost of the lattice. The density profile evolution has an intermediate state just like the global low density phase steady state, which is indicated by green color ($t = 6000$) in Figure 2.9. After the intermediated state, the density profile continues evolving until the global coexistence line steady state is reached.

We divide the transient state of the coexistence line into three stages: $[0, L]$, $(L, t_c^{CL}]$ and $(t_c^{CL}, T_c^{CL}]$, where t_c^{CL} is the time that the lattice reaches the intermediate state (the global low density steady state) and T_c^{CL} is the time that the lattice reaches the global coexistence line steady state. The leftmost part of the lattice is treated as a sub system with a particle entering rate $\alpha < \frac{1}{2}$ and a maximal particle exit rate, 1. The

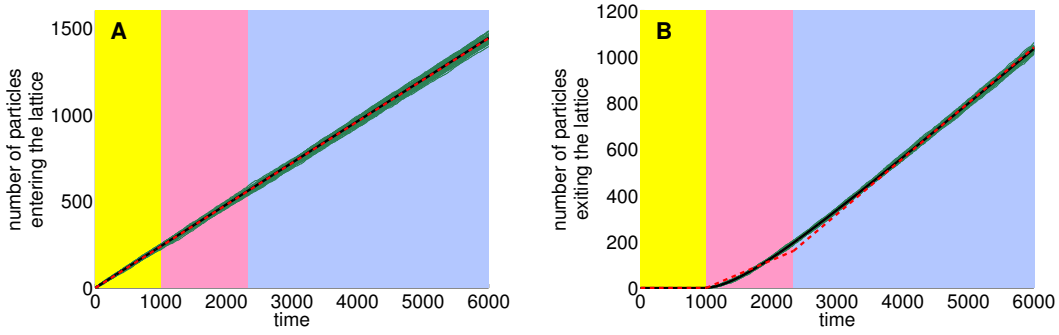


Figure 2.10: **Transient state approximation for coexistence line.** (A) Number of particles entering the lattice and (B) number of particles exiting lattice are plotted as functions of time. Red dashed lines are approximations (2.3.8) and black lines are stochastic averages from 2000 simulations. 20 realizations of simulation are plotted (green). Yellow region is the first stage and pink region is the second stage. In purple region, TASEP is in the steady state. We use $\alpha = 0.4$, $\beta = 0.4$, $L = 1000$.

sub system is in the low density phase. Therefore, during all the three stages, $[0, T_c^{CL}]$, J_1 is determined by the current of the sub system, which is approximately given by $\alpha(1-\alpha)$. In the first stage, no particle leaves, so $J_2 = 0$. In the second stage, $(L, t_c^{CL}]$, particles start leaving the lattice, thus J_2 increases from 0 until the intermediate steady state $\alpha(1-\alpha)$ is reached. As an assumption, we set $J_2 = \dfrac{\alpha(1-\alpha)}{2}$ during the second stage. In the third stage $(t_c^{CL}, T_c^{CL}]$, though the global steady state is yet to be reached, the leftmost and the rightmost of the lattice are approximately already in the steady state, so $J_1 = J_2 = \alpha(1-\alpha)$. After T_c^{CL}, the whole lattice is in the global coexistence line steady state, $J_1 = J_2 = \alpha(1-\alpha)$. In this sense, it is insignificant to estimate T_c^{CL}. In fact, T_c^{CL} is much larger than t_c^{CL} and is hard to be estimated. This is not difficult to understand. Because during the $(t_c^{CL}, T_c^{CL}]$, though $J_1 = J_2$ by mean field approximation, J_1 and J_2 have a slight stochastic difference which can be positive or negative (overall it is positive). The density profile evolves with this slight stochastic difference and takes a long period of time. At t_c^{CL}, the lattice reaches the intermediate state just like the global steady state of the low density phase. The number of particles on the lattice at time t_c^{CL} is approximately given by the number of particles on the lattice at the low density phase steady state, and also given by the net number of particles

added to the lattice during $[0, t_c^{CL}]$:

$$\alpha(1-\alpha)L + \left(\alpha(1-\alpha) - \frac{\alpha(1-\alpha)}{2}\right)\left(t_c^{LD} - L\right) = \alpha L \quad \Longrightarrow \quad t_c^{CL} = \frac{1+\alpha}{1-\alpha}L.$$

The red term is net number of particles added during the first stage $[0, L]$, and the blue term is the net number of particles added to the lattice during the second stage. The green term, αL, is the number of particles on the lattice in the intermediate state, which is given by the product of average density in the intermediate state and the length of the lattice. In summary, in the coexistence line:

$$J_1 = \alpha(1-\alpha), \quad t > 0; \qquad J_2 = \begin{cases} 0, & 0 < t < L, \\ \dfrac{\alpha(1-\alpha)}{2}, & L \leqslant t < t_c^{LD}, \\ \alpha(1-\alpha), & t \geqslant t_c^{LD}, \end{cases} \tag{2.3.7}$$

and

$$N_1 = \alpha(1-\alpha)t; \quad N_2 = \begin{cases} 0, & 0 < t < L, \\ \dfrac{\alpha(1-\alpha)}{2}(t-L), & L \leqslant t < t_c^{CL}, \\ \alpha(1-\alpha)t - \alpha L, & t \geqslant t_c^{CL}. \end{cases} \tag{2.3.8}$$

The transient behaviour approximation for the coexistence line is then given by (2.3.7), (2.3.8), and is shown in Figure 2.10. We can see from the figure that the approximation (2.3.8) agrees well with the average of simulations, though not perfect. This is because we used an approximated intermediate state.

2.3.4 Maximal Current Phase

TASEP is in *maximal current phase* if $\alpha, \beta \geqslant \frac{1}{2}$. A typical time evolution of density profile of the maximal current phase is shown in Figure 2.11. We find that the density

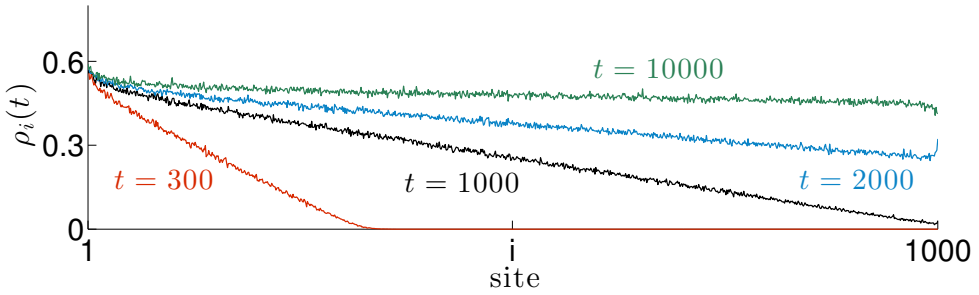


Figure 2.11: **Time evolution of density profile for maximal current phase.** Density profile $\rho_i(t)$ is plotted, different colours show different times. Each of $\rho_i(t)$ is the average of 2000 simulations. We use $\alpha = 0.6, \beta = 0.6, L = 1000$.

profiles at different times are approximately linear and intersect at the leftmost of the lattice. We divide the transient state into two stages: $[0, L]$ and $(L, t_c^{MC}]$, where t_c^{MC} is the time that the whole lattice is approximately in a global steady state. The leftmost part of the lattice is treated as a sub system with a particle entering rate $\alpha \geqslant \frac{1}{2}$ and a maximal particle exit rate, 1. The sub system is in the maximal current phase. Therefore, during $[0, t_c^{MC}]$, J_1 is determined by the current of the sub system, which is approximately given by $\frac{1}{4}$. In the first stage, no particle leaves, so $J_2 = 0$. In the second stage, particles start leaving the lattice, thus J_1 increases from 0 until the steady state $\frac{1}{4}$ is reached. As an assumption, we set $J_2 = \frac{1}{8}$ during the second stage. After t_c^{MC}, the whole lattice is in the global maximal current phase steady state, $J_1 = J_2 = \frac{1}{4}$.

At t_c^{MC}, the net number of particles added to the lattice in $[0, t_c^{MC}]$ can be approximated by the total number of particles on the lattice in the steady state:

$$\frac{1}{4}L + \left(\frac{1}{4} - \frac{1}{8}\right)\left(t_c^{MC} - L\right) = \frac{1}{2}L \qquad \Longrightarrow \qquad t_c^{MC} = 3L,$$

The red term is the net number of particles added to the lattice during $[0, L]$, and the blue term is the net number of particles added to the lattice during $(L, t_c^{MC}]$. The green term, $\frac{1}{2}L$, is the the number of particles on the lattice in the steady state, which is given by product of average density in the global steady state and the length of the lattice. In

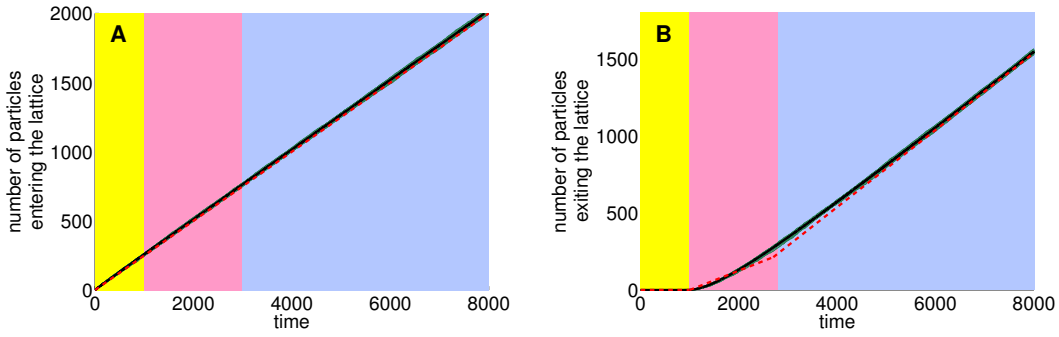


Figure 2.12: **Transient state approximation for maximal current phase.** (A) Number of particles entering the lattice and (B) number of particles exiting lattice are plotted as functions of time. Red dashed lines are approximations (2.3.10) and black lines are stochastic averages from 2000 simulations. 20 realizations of simulation are plotted (green). Yellow region is the first stage and pink region is the second stage. In purple region, TASEP is in the steady state. We use $\alpha = 0.6$, $\beta = 0.6$, $L = 1000$.

summary, in the maximal current phase:

$$J_1 = \frac{1}{4}, \quad t > 0; \qquad J_2 = \begin{cases} 0, & 0 < t < L, \\ \frac{1}{8}, & L \leqslant t < t_c^{MC}, \\ \frac{1}{4}, & t \geqslant t_c^{MC}. \end{cases} \tag{2.3.9}$$

The total numbers of particles entering and exiting the lattice are:

$$N_1 = \frac{1}{4}t, \quad t > 0; \qquad N_2 = \begin{cases} 0, & 0 < t < L, \\ \frac{1}{8}(t-L), & L \leqslant t < t_c^{MC}, \\ \frac{1}{4}t - \frac{1}{2}L, & t \geqslant t_c^{MC}. \end{cases} \tag{2.3.10}$$

The transient state approximation for the maximal current phase is then given by (2.3.9), (2.3.10), and is shown in Figure 2.12. The approximation (2.3.10) agrees well with the average of simulations, though not perfect. Nonlinear transient state need to be used to improve the approximation.

2.4 Simulation Methods

In this section, we will introduce two main simulation methods for TASEP: Gillespie Stochastic Simulation Algorithm (Gillespie SSA) and Monte Carlo method.

2.4.1 Gillespie SSA

We first introduce the Gillespie SSA. $\boldsymbol{S}(t) = [S_1(t), \ldots, S_L(t)]^T$ is called a *state vector* for the TASEP. The state vector describes the state of TASEP at time t. In TASEP, there are $L+1$ reactions that can take place. Let $\boldsymbol{R} = [R_0, \ldots R_L]^T$ represent these reactions. R_0 is the reaction that a particle binds to the empty site 1 with rate α, R_L is the reaction that the particle on site L leaves the lattice with rate β, and R_j, $j \in [1, L-1]$ is the reaction that the particle on site j jumps to the empty site $j+1$. The reactions are associated with so called *propensity functions*, $\boldsymbol{a}(t) = [a_0(t), \ldots, a_L(t)]^T$. The probability that the reaction R_j takes place in an infinitesimal time interval $[t, t+dt)$ is given by $a_j(t)dt$. The propensity functions for TASEP is given by:

$$a_0(t) = \alpha(1 - S_1(t)), \quad a_L(t) = \beta S_L(t), \quad a_i(t) = S_i(t)(1 - S_{i+1}(t)), i \in [1, L-1]. \tag{2.4.1}$$

The essence of the Gillespie SSA is to determine which reaction is going to take place and when it is going to happen. Next, we give a brief introduction to how the algorithm determines when the specific next reaction takes place.

Consider the probability that no reactions occur during $[t, t+\tau)$ given a state $\boldsymbol{S}(t)$, this probability is denoted by $P_0\left(\tau \,|\, \boldsymbol{S}(t), t\right)$. Under common assumption (e.g., large number of reactions), and given the Markovian dynamics, reactions that take place over $[t, t+\tau)$ are independent of those in $[t+\tau, t+\tau+d\tau)$. Then during $[t, t+\tau+d\tau)$,

prob. no reactions occur over $[t, t+\tau+d\tau)$

= prob. no reactions occur over $[t, t+\tau)\times$ prob. no reactions occur over $[t+\tau, t+\tau+d\tau)$

= prob. no reactions occur over $[t, t+\tau) \times (1 -$ sum of prob. each reaction occurs over $[t+\tau, t+\tau+d\tau)\big)$,

which yields

$$P_0\left(\tau+d\tau \mid \boldsymbol{S}(t), t\right) = P_0\left(\tau \mid \boldsymbol{S}(t), t\right)\left(1-\sum_{j=0}^{L} a_j(t+\tau)d\tau\right)$$

Since we consider the situation that no reactions occur over $[t, t+\tau+d\tau)$, so $\boldsymbol{S}(t+\tau) = \boldsymbol{S}(t)$. That is,

$$P_0\left(\tau+d\tau \mid \boldsymbol{S}(t), t\right) = P_0\left(\tau \mid \boldsymbol{S}(t), t\right)\left(1-\sum_{j=0}^{L} a_j(t)d\tau\right)$$

Subtracting $P_0\left(\tau \mid \boldsymbol{S}(t), t\right)$ from both sides and dividing by $d\tau$, then letting $d\tau \to 0$, lead to the following ODE,

$$\frac{dP_0\left(\tau \mid \boldsymbol{S}(t), t\right)}{d\tau} = -a_{sum}(t)P_0\left(\tau \mid \boldsymbol{S}(t), t\right), \qquad a_{sum}(t) = \sum_{j=0}^{L} a_j(t).$$

Solving this ODE, one obtains

$$P_0\left(\tau \mid \boldsymbol{S}(t), t\right) = e^{-a_{sum}(t)\tau}. \tag{2.4.2}$$

Consider the probability that the reaction R_j occurs in the time interval $[t+\tau, t+\tau+d\tau)$, which is given by $p\left(\tau, j \mid \boldsymbol{S}(t), t\right)d\tau$. Note that $d\tau$ is so small that there can reasonably be only one reaction occurring over $[t+\tau, t+\tau+d\tau)$. Because:

$p\left(\tau, j \mid \boldsymbol{S}(t), t\right)d\tau$

= prob. no reaction occur over $[t, t+\tau)\times$ prob. reaction R_j occurs over $[t+\tau, t+$

$$\tau + d\tau)$$
$$= P_0\left(\tau \mid \boldsymbol{S}(t), t\right) a_j(t) d\tau = a_j(t) e^{-a_{sum}(t)\tau} d\tau,$$

so

$$p\left(\tau, j \mid \boldsymbol{S}(t), t\right) = a_j(t) e^{-a_{sum}(t)\tau} = \frac{a_j(t)}{a_{sum}(t)} a_{sum}(t) e^{-a_{sum}(t)\tau}. \tag{2.4.3}$$

The fraction $\dfrac{a_j(t)}{a_{sum}(t)}$ is used to determine the reaction index j. $(0, 1)$ interval can be divided into $L+1$ subintervals, whose jth subinterval has a length of $\dfrac{a_j(t)}{a_{sum}(t)}$. Therefore the chance of the reaction R_j being chosen is proportional to its propensity function. Draw a uniform $(0, 1)$ random number ξ_1, then j is chosen by:

$$\sum_{k=0}^{j-1} a_k(t) < \xi_1 a_{sum}(t) \leqslant \sum_{k=0}^{j} a_k(t). \tag{2.4.4}$$

Notice that $a_{sum}(t) e^{-a_{sum}(t)\tau}$ is an exponential distribution, which describes the time between events in a Poisson process. For any given $0 < b < c$:

$$\mathbb{P}(b < \tau < c) = \int_b^c a_{sum}(t) e^{-a_{sum}(t)\tau} d\tau = e^{-a_{sum}(t)b} - e^{a_{sum}(t)c},$$

$$\mathbb{P}(b < \tau < c) = \mathbb{P}(-c < -\tau < -b) = \mathbb{P}\left(e^{-a_{sum}(t)c} < e^{-a_{sum}(t)\tau} < e^{-a_{sum}(t)b}\right).$$

Therefore,

$$\mathbb{P}\left(e^{-a_{sum}(t)c} < e^{-a_{sum}(t)\tau} < e^{-a_{sum}(t)b}\right) = e^{-a_{sum}(t)b} - e^{a_{sum}(t)c},$$

which indicates that $e^{-a_{sum}(t)\tau}$ is a uniform distribution. Since $0 < e^{-a_{sum}(t)\tau} < 1$, so draw a uniform $(0, 1)$ random number ξ_2, τ is determined by solving $e^{-a_{sum}(t)\tau} = \xi_2$:

$$\tau = \frac{1}{a_{sum}(t)} \ln\left(\frac{1}{\xi_2}\right). \tag{2.4.5}$$

The algorithm is summarised below.

1. Initialize TASEP: $S_i = 0, i \in [1, L]$, $t = 0$ and $a_0 = \alpha$, $a_i = 0, 0 < i \leqslant L$, $a_{sum} = \alpha$,
2. Draw a uniform random number ξ_1, obtain τ according to (2.4.5), and set $t = t + \tau$,
3. Draw a uniform random number ξ_2, find j according to (2.4.4):

   ①: If $j = 0$, set $S_1 = 1, a_0 = 0, a_1 = 1 - S_2, a_{sum} = a_{sum} - \alpha + 1 - S_2$;

   ②: If $j = 1$, set $S_1 = 0, S_2 = 1, a_0 = \alpha, a_1 = 0, a_2 = 1 - S_3, a_{sum} = a_{sum} + \alpha - S_3$;

   ③: If $j = i, i \in [2, L-2]$, set $S_i = 0, S_{i+1} = 1, a_{i-1} = S_{i-1}, a_i = 0, a_{i+1} = 1 - S_{i+2}, a_{sum} = a_{sum} + S_{i-1} - S_{i+2}$;

   ④: If $j = L - 1$, set $S_{L-1} = 0, S_L = 1, a_{L-2} = S_{L-2}, a_{L-1} = 0, a_L = \beta$ and $a_{sum} = a_{sum} - 1 + S_{L-2} + \beta$;

   ⑤: If $j = L$, set $S_L = 0, a_{L-1} = S_{L-1}, a_L = 0, a_{sum} = a_{sum} + S_{L-1} - \beta$,

4. Repeat steps $2 - 3$ or else stop the process after some preset time steps.

2.4.2 Monte Carlo Method

An alternative simulation algorithm is the Monte Carlo method. To apply the Monte Carlo method to simulate the TASEP, a time step, which is called a Monte Carlo step (MCS) is defined first. The lattice has L sites labelled $1, \ldots, L$. A particle reservoir is indicated by site 0. A MCS is defined as $L + 1$ 'attempts'. At each 'attempt', a site $i, i \in [0, L]$ is randomly chosen, and its occupation S_i is updated. On the average, all the sites are chosen once in a MCS. The algorithm is summarised below.

1. Initialize TASEP: $S_i = 0, i \in [1, L]$ and $t = 0$, set a MCS, τ,

2. Randomly choose a site:

   ①: If site 0 is chosen and $S_1 = 0$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \alpha\tau$, set $S_1 = 1$;

   ②: If site i, $i \in [1, L-1]$ is chosen, and $S_i = 1$, $S_{i+1} = 0$, set $S_i = 0$, $S_{i+1} = 1$;

   ③: If site L is chosen and $S_L = 1$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \beta\tau$, set $S_L = 0$,

3. Repeat step 2 for $L+1$ times, then set $t = t + \tau$,
4. Repeat step $2-3$ or else stop the process after some preset time steps.

Both the Monte Carlo method and the Gillespie SSA can be used to generate stochastic simulations for TASEP. The Monte Carlo method randomly chooses a reaction and determines the reaction to occur or not. The Gillespie SSA randomly determine when and which reaction to occur next. The time step of Monte Carlo method is fixed while the time step of Gillespie SSA is random. Therefore, it is easier to average realizations of simulations by Monte Carlo method. Moreover, it is much faster to run the simulations using the Monte Carlo method than the Gillespie SSA without further optimizing the SSA algorithm. We took a test using C programming, to run one realization of simulation, the Gillespie SSA took 4.957864 seconds, whilst the Monte Carlo method took 0.138277 seconds. Thus we will use Monte Carlo method in this thesis unless otherwise noted. The algorithm is displayed in Appendix C

2.5 Summary

We outlined the exact steady state solution and the mean field solution of the TASEP which were originally constructed by Derrida *et al*. Compared to the original results,

Low density phase	$\alpha < \beta, \alpha < 1/2$	$\rho = \alpha$	$J = \alpha(1-\alpha)$
High density phase	$\alpha > \beta, \beta < 1/2$	$\rho = 1-\beta$	$J = \beta(1-\beta)$
Coexistence line phase	$\alpha = \beta < 1/2$	$\rho = 1/2$	$J = \alpha(1-\alpha)$
Maximal current phase	$\alpha, \beta \geqslant 1/2$	$\rho = 1/2$	$J = 1/4$

Table 2.1: **Summary of the mean field solutions for TASEP.** Steady state for $L \to \infty$ is shown. γ is set to be 1.

we filled in some steps to provide the reader with a fairly rigorous and more comprehensive description of the results construction. The exact steady state solution was shown in Lemma 2.4. The solution is valid for systems of arbitrary size L. One can obtain the exact expression for current, average occupation in the steady state.

We also displayed the mean field steady state solution for system of large size. Four phases were defined according to different values of α and β. These phases were named low density, high density, coexistence (shock phase) and maximal current. The mean field solution in each phase is summarised in Table 2.1.

We also constructed approximations for the transient state of the TASEP before the steady state, which may be important for some applications, e.g., application of translation process modelling. Proteins are made throughout the mRNA lifetime. For modelling of translation of those mRNAs with short lifetime, the transient state of the TASEP can take up fair share of the mRNA lifetime. Thus only considering the steady state of the TASEP is not comprehensive. We gave linear transient state approximation of the rates of particles entering and exiting the lattice in Section 2.3.

We introduced two stochastic simulation methods for TASEP. The Gillespie SSA and the Monte Carlo method. Unless otherwise noted, we will use Monte Carlo method in this thesis. Our aim of this chapter was to give an overview of the fundamental stochastic particle interacting model, the TASEP, which is the basis of the whole thesis. Later, we will apply the TASEP model to study how protein is made by the mRNA translation process and how protein production is controlled.

Chapter 3

The Role of mRNA Degradation in Mediating Translation

3.1 Introduction

In this chapter, we study an incorporated process of prokaryotic mRNA degradation-translation and how this affects protein production. By developing a TASEP based model and mean field approximation, we are able to obtain the probability density distribution of protein production, and further to give some strategies to control protein production from the mathematical point of view. As we stated in the Introduction chapter, the degradation of prokaryotic mRNAs is usually initiated by a simple single step: binding of a *degradosome* to a target mRNA. The degradosome is a multi-protein complex present in most prokaryotic cells [21]. A degradosome binds to the 5' end of an mRNA at a special site upstream of the ribosome binding site. Subsequently, it cleaves the ribosome binding site, initiating the mRNA degradation. There are two main degradation types: *tracking degradation* and *rapid degradation*.

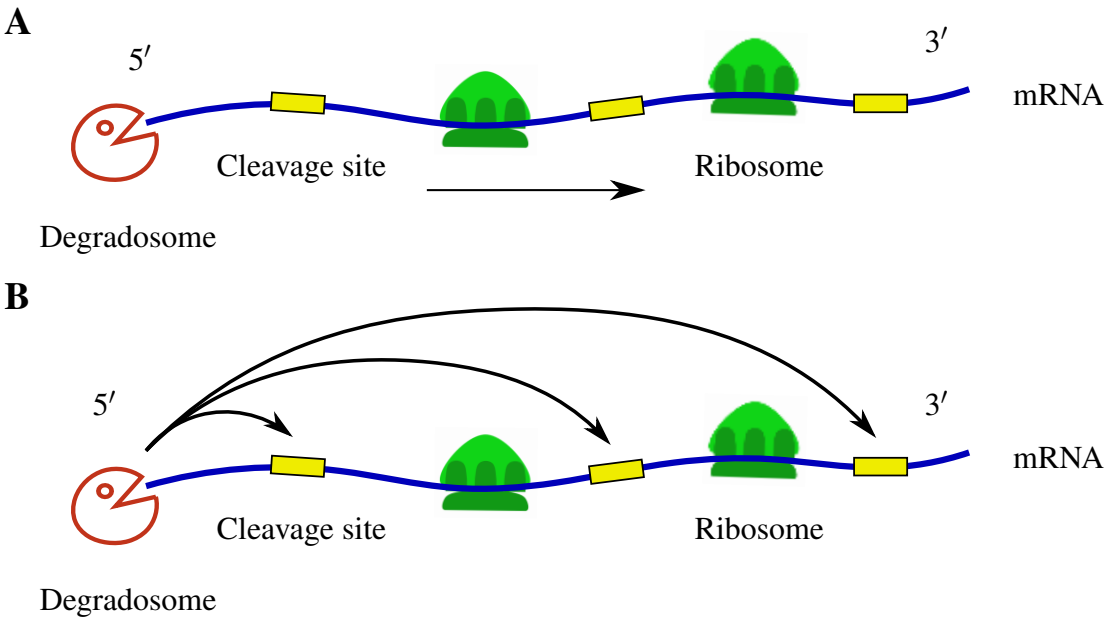


Figure 3.1: **Schematic of mRNA degradation.** A degradosome binds to the 5' end of an mRNA, (A) tracks along the mRNA to find cleavage sites in 5'-3' direction or (B) loops to cleavage sites or accesses to cleavage sites directly. The mRNA is cleaved at those cleavage sites. Type A is called tracking degradation. Type B is called rapid degradation.

The schematic of the tracking degradation is shown in Figure 3.1A. The degradosome cleaves the mRNA at certain specific sites. Those specific sites are called *cleavage sites*. Cleavage sites are specific codons within the mRNA to which the degradosome can bind. Once bound, the degradosome cleaves nucleotides by hydrolysing the phosphodiester bond between them. In tracking degradation, the degradosome cannot access all cleavage sites directly. Rather, the degradosome must track along 5'-3' direction sequentially cleaving the mRNA at the cleavage sites. One example of the tracking degradation is the mRNA degradation in *B. subtilis* [10, 28]. In the tracking degradation, ongoing translation can inhibit degradation because ribosomes on the mRNA block the degradosome from scanning the downstream cleavage sites. In other words, after an mRNA in active translation being triggered to decay by a degradosome, the ribosomes downstream of the degradosome are able to finish translation processes.

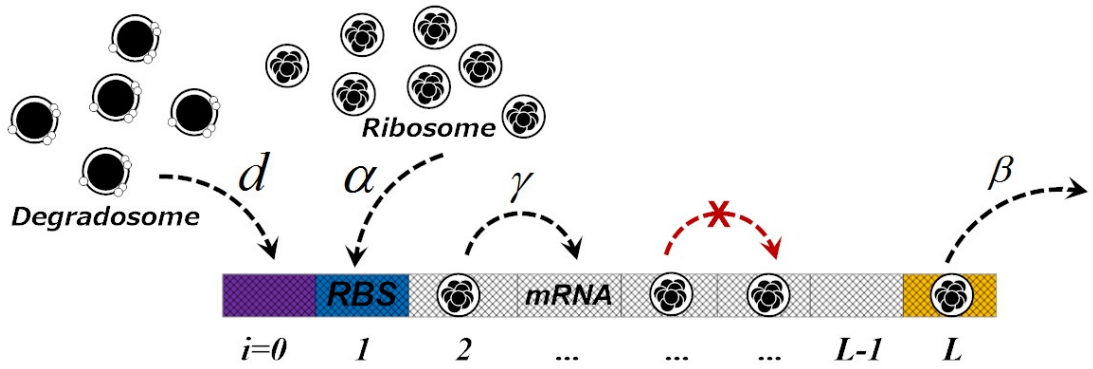


Figure 3.2: **Illustration of mRNA degradation-translation model.** A degradosome binds to site 0 to initiate mRNA degradation. Degradosome binding is independent of ribosome binding. Degradosome binding rate is d. Ribosome binds to site 1 at rate α, hops to next empty site at rate γ and leaves site L at rate β.

The schematic of the rapid degradation is shown in Figure 3.1B. In rapid degradation, the degradosome can loop to the cleavage sites or even access to the cleavage sites directly to cut the mRNA into pieces once after the initiation of the mRNA degradation. Thus, the mRNA is degraded almost immediately. One example of the rapid degradation is the mRNA degradation in *E. coli* [22, 28, 82].

Synthesis and maturation of eukaryotic mRNA occur exclusively in the nucleus. Only after the mRNA is exported to the cytoplasm, translation and degradation can occur. In contrast to eukaryotes, prokaryotic mRNA is transcribed, translated, and degraded simultaneously. The 5' end of the mRNA starts to decay even before the 3' end has been synthesised or translated [94]. In this sense, we simply model the initiation of prokaryotic mRNA degradation to a single step, degradosome binding (to include all the possible biochemical interactions). We now construct an mRNA degradation - translation model based on the TASEP framework. We illustrate this model in Figure 3.2. A degradosome binds to site 0 (the degradosome binding site) at rate d to initiate mRNA degradation. Ribosome binds to the empty ribosome binding site at rate α. Ribosomes on site $i, i \in [1, L-1]$ jump to site $i+1$ at rate γ provided site $i+1$ is empty. Ribosome leaves site L at rate β. Note that for simulation, γ has been chosen

to be one, unless otherwise noted.

As above we use Monte Carlo method to simulate the mRNA degradation - translation model. We introduced the Monte Carlo simulation method in Section 2.4.2. To simulate the mRNA degradation - translation model, a Monte Carlo step consists of $L+2$ attempts (to include randomly choosing the degradosome particle binding site). If the degradosome binding site is chosen and it is empty, then a uniform random number $\xi \in [0,\ 1]$ is drawn. If $\xi < d\tau$, then the degradosome binding site is occupied and the system can no longer load ribosomes. The scheme afterwards depends on the degradation type. For the tracking degradation, those ribosome particles already on the lattice continue moving until the whole lattice is empty. For the rapid degradation, the simulation stops when degradosome particle binding occurs. To measure the amount of protein produced from one mRNA during its lifetime, the number of ribosome particles that leave site L during simulation are counted.

3.2 Degradosome Binding Can be Modelled by a Poisson Process

Both the tracking degradation and the rapid degradation are initiated by degradosome binding. We assumed that degradosome binds to the 5' end of an mRNA at a constant rate d, thus the degradosome binding can be modelled as a Poisson process. Thus, the waiting time to degradosome binding is exponentially distributed. Exponentially distributed turnover of the mRNAs has been found and used [41, 108, 129, 135]. Let T be a random variable representing the degradosome binding time and $f_T(t)$ be the probability density function for T. We are able to write down $f_T(t)$ directly:

$$f_T(t) = de^{-dt}, \tag{3.2.1}$$

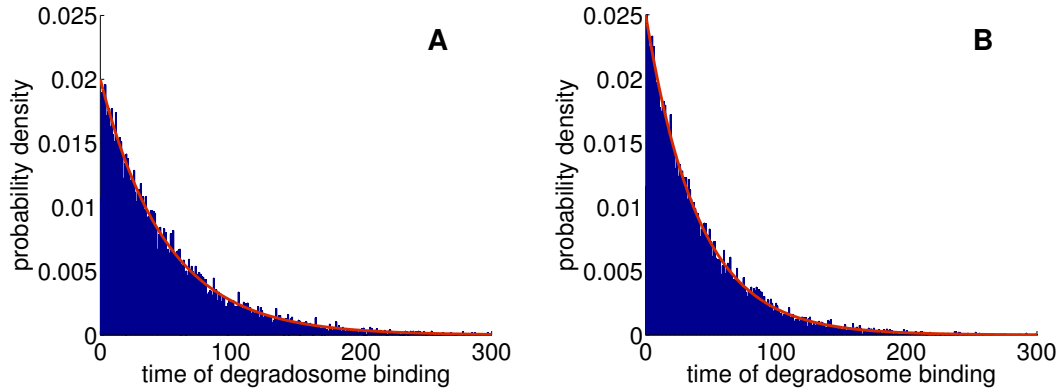


Figure 3.3: **Probability densities of degradosome binding time for different degradosome binding rates.** Red lines are probability density function given by formula (3.2.1), and blue bars are results from Monte Carlo simulations (please refer to text for details). $L = 1000, \alpha = 0.3, \beta = 0.1$. (A) $d = 0.02$. (B) $d = 0.025$.

whose mean value and standard deviation are both $1/d$. We used the simulation method introduced in Section 3.1 to run 15000 simulations. We recorded the waiting time to degradosome particle binding and obtained corresponding probability density distributions. The probability density distributions are shown in Figure 3.3. The simulation results (blue histograms) match the theoretic formula (3.2.1) very well (e.g., the exact mean value and standard deviation for $d = 0.02$ are both 50, whilst the mean value of the stochastic simulation is 49.9698 and the standard deviation is 50.2522 to 4 decimal places). Thus, the Monte Carlo simulation is consistent with the model set up.

3.3 Protein Distribution under Tracking Degradation

The total number of proteins produced from one mRNA is affected by mRNA degradation. Intuitively thinking, the longer mRNA lifetime, the more protein is produced. As an assumption, mRNA degradation is initiated by a single step: degradosome binding, and the degradosome binding time is exponentially distributed. For the tracking degradation, at the time of degradosome binding, all the ribosomes bound onto the

mRNA finish translation to produce proteins. Therefore, the number of proteins produced from one mRNA is equivalent to the number of ribosomes that bind to the RBS before degradosome binding occurs. We now use the method described in Section 3.1 to generate simulations.

Recall that in Section 2.3, we approximated the average number of ribosome particles entering the lattice by N_1, where N_1 is a function of time, t. In fact, by replacing t with T (T is the random variable representing degradosome binding time), N_1 can be used to approximate the expected number of proteins, N, per mRNA for the tracking degradation.

Before moving to the next step, we recall that:

Lemma 3.1. *If X is a random variable whose probability density function (PDF) is $f_X(x)$, then the random variable $Y = aX + b$, where a, b are constants and $a > 0$, has PDF $f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$.*

3.3.1 Low Density Phase

Recall from (2.3.2) that in the low density phase, N_1 is approximately given by $N_1 = \alpha(1-\alpha)T$. The probability density function of T is given by (3.2.1). Applying Lemma 3.1, we can write down the probability density function for protein number $N = n$ directly:

$$f_N(n) = \frac{d}{\alpha(1-\alpha)} \exp\left\{\frac{-dn}{\alpha(1-\alpha)}\right\}, \qquad n \in \mathbb{R}. \tag{3.3.1}$$

Probability density function (3.3.1) is an exponential distribution, whose expected value and standard deviation are both $\alpha(1-\alpha)/d$. We ran 50000 simulations using the method introduced in Section 3.1. We recorded number of proteins produced

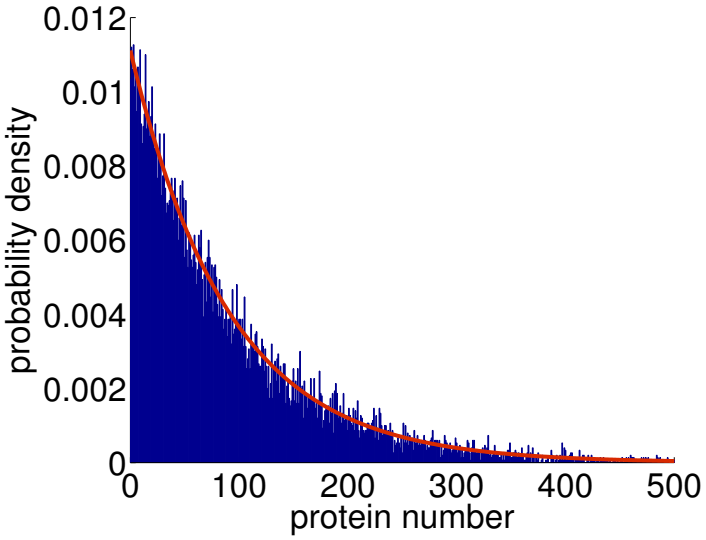


Figure 3.4: **Probability density distribution of protein for low density phase: tracking degradation.** Red curve is probability density function (3.3.1). Blue bars are probability density distribution from simulations (please refer to the text for details). $\alpha = 0.1, \beta = 0.8, L = 1000, d = 0.001$.

in each realization of simulation and obtained corresponding probability density distribution. The probability density distribution of protein is shown in Figure 3.4. Simulation and the approximated probability density function are in good agreement (the mean value of simulation is 89.8374 and the standard deviation is 90.1164, whilst the expected value and the standard deviation of (3.3.1) are both 90). We see that increasing the ribosome binding rate, α, or decreasing the degradosome binding rate, d, (or both) contribute to a higher expected value and standard deviation, which are given by $\frac{\alpha(1-\alpha)}{d}$. This is very straightforward from a biological point of view, larger α leads to more ribosomes participating in translation and smaller d prolongs mRNA lifetime. Note, however, that the probability density function of protein (3.3.1) is independent of L and β. This is because in the low density phase, changing β does not change any properties of the TASEP. Moreover, for tracking mRNA degradation, no matter how long the mRNA is, ribosomes which bind to the mRNA before degradosome binding occurs always finish translation.

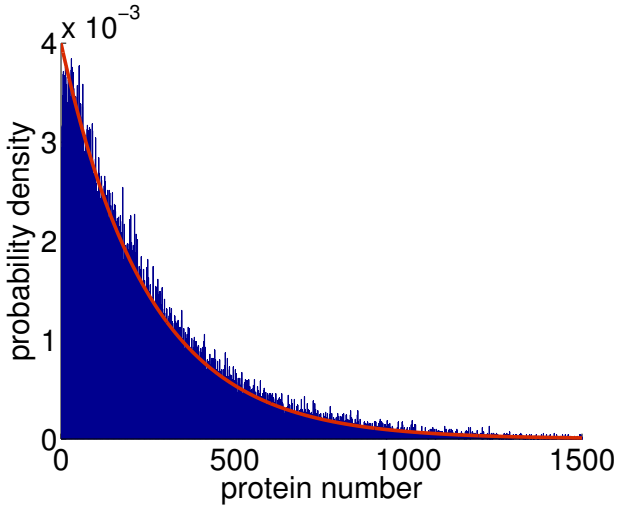


Figure 3.5: **Probability density distribution of protein for maximal current phase: tracking degradation.** Red curve is probability density function (3.3.2). Blue bars are probability density distribution from simulations (please refer to the text for details). $\alpha = 0.6, \beta = 0.6, L = 1000, d = 0.001$.

3.3.2 Maximal Current Phase

Recall from (2.3.10) that in the maximal current phase, N is approximately given by $N = \frac{1}{4}T$. The probability density function for protein number $N = n$ can be simply derived:

$$f_N(n) = 4d \exp\{-4dn\}, \qquad n \in \mathbb{R}. \tag{3.3.2}$$

Probability density function (3.3.2) is an exponential distribution, whose expected value and standard deviation are both $\frac{1}{4d}$. Smaller d gives a higher expected value and standard deviation. (3.3.2) is only dependent on d, and is independent of α, β, L. This is because in the maximal current phase, α and β do not affect any properties of the TASEP. Moreover, for tracking mRNA degradation, ribosomes always finish translation no matter how long the mRNA is. We run 50000 simulations using the method introduced in Section 3.1. The result is shown in Figure 3.5. Our simulation and the approximated probability density function are in good agreement. A small systematic error arises because that the current in simulation is actually slightly larger than $\frac{1}{4}$.

3.3.3 High Density Phase: $\alpha < 1/2$

Recall from Section 2.3 that transient behaviours in the high density phase with $\alpha < \frac{1}{2}$ and $\alpha \geqslant \frac{1}{2}$ are different. Thus, the protein production from one mRNA can also be divided into two cases. When $\alpha < \frac{1}{2}$, N is approximately given by (see (2.3.4)):

$$N = \begin{cases} \alpha(1-\alpha)T, & T < t_{c1}^{HD}, \\ \beta(1-\beta)T + (1-\beta)^2 L, & T \geqslant t_{c1}^{HD}, \end{cases} \qquad t_{c1}^{HD} = \frac{(1-\beta)^2 L}{\alpha(1-\alpha) - \beta(1-\beta)}.$$

The probability density function for protein number $N = n$ can be derived using Lemma 3.1:

$$f_N(n) = \begin{cases} \dfrac{d}{\alpha(1-\alpha)} \exp\left\{ \dfrac{-dn}{\alpha(1-\alpha)} \right\}, & n < \alpha(1-\alpha)t_{c1}^{HD}, \\ \dfrac{d}{\beta(1-\beta)} \exp\left\{ -d\left(\dfrac{n}{\beta(1-\beta)} - \dfrac{1-\beta}{\beta} L \right) \right\}, & n \geqslant \alpha(1-\alpha)t_{c1}^{HD}. \end{cases} \tag{3.3.3}$$

From Figure 3.6, we see that probability density function (3.3.3) has two branches. One branch is an exponential distribution with parameter $\dfrac{d}{\alpha(1-\alpha)}$, which can be regarded as a distribution of protein being produced at rate $\alpha(1-\alpha)$. The other branch can be obtained by shifting an exponential distribution with parameter $\dfrac{d}{\beta(1-\beta)}$, which can be regarded as a distribution of protein being produced at rate $\beta(1-\beta)$. Moreover, we notice that the probability density function is discontinuous at a critical point. This is because the protein production rate at the critical point changes to $\beta(1-\beta)$ from $\alpha(1-\alpha)$ (see Section 2.3.2.1). We will study the properties of the critical point later.

Let $E[N]$ be the expected value of (3.3.3), then $E[N]$ is given by:

$$E[N] = \int_0^\infty n f_N(n) dn.$$

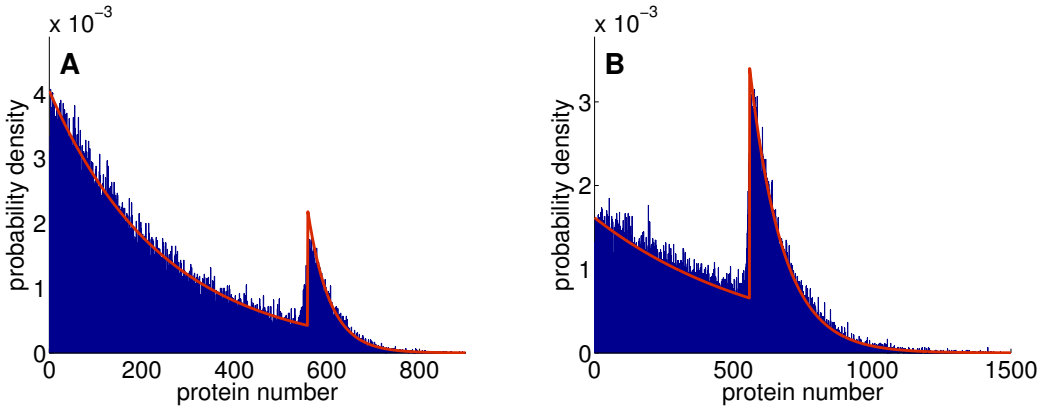


Figure 3.6: **Probability density distribution of protein for high density phase under tracking degradation:** $\alpha < 1/2$. Red curve is probability density function (3.3.3). Blue bars are probability density distribution from simulations (please refer to the text for details). (A) $\alpha = 0.45, \beta = 0.05, L = 500, d = 0.001$. (B) $\alpha = 0.45, \beta = 0.05, L = 500, d = 0.0004$.

We obtain:

$$\begin{aligned} E[N] &= \int_0^\infty n f_N(n) dn \\ &= \int_0^{\alpha(1-\alpha)t_{c1}^{HD}} n \frac{d}{\alpha(1-\alpha)} \exp\left\{\frac{-dn}{\alpha(1-\alpha)}\right\} dn \\ &\quad + \int_{\alpha(1-\alpha)t_{c1}^{HD}}^\infty n \frac{d}{\beta(1-\beta)} \exp\left\{-d\left(\frac{n}{\beta(1-\beta)} - \frac{1-\beta}{\beta}L\right)\right\} dn \\ &= \frac{\alpha(1-\alpha)}{d} + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d} \exp\left\{-dt_{c1}^{HD}\right\}. \end{aligned} \tag{3.3.4}$$

To study the properties of $E[N]$, we take derivatives of $E[N]$ with respect to α, β, d, L, respectively:

$$\begin{aligned} \frac{\partial E[N]}{\partial \alpha} &= \frac{1-2\alpha}{d} + \frac{-1+2\alpha}{d} \exp\left\{-dt_{c1}^{HD}\right\} \\ &\quad + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d} \exp\left\{-dt_{c1}^{HD}\right\} \frac{d(1-\beta)^2 L(1-2\alpha)}{\left(\alpha(1-\alpha)-\beta(1-\beta)\right)^2} \\ &= \frac{1-2\alpha}{d} + \frac{-1+2\alpha}{d} \exp\left\{-dt_{c1}^{HD}\right\} + \exp\left\{-dt_{c1}^{HD}\right\} \frac{(1-\beta)^2 L(1-2\alpha)}{\beta(1-\beta)-\alpha(1-\alpha)} \end{aligned}$$

$$= \frac{1-2\alpha}{d}\exp\left\{-dt_{c1}^{HD}\right\}\left(\exp\left\{dt_{c1}^{HD}\right\} - \left(1+\frac{d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right)\right)$$

$$> \frac{1-2\alpha}{d}\exp\left\{-dt_{c1}^{HD}\right\}\left(1+dt_{c1}^{HD} - \left(1+\frac{d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right)\right)$$

$$= \frac{1-2\alpha}{d}\exp\left\{-dt_{c1}^{HD}\right\}\left(1+dt_{c1}^{HD} - \left(1+dt_{c1}^{HD}\right)\right) > 0$$

$$\because \alpha < 1/2 \Longrightarrow 1-2\alpha > 0,$$

$$\frac{\partial E[N]}{\partial \beta} = \frac{1-2\beta}{d}\exp\left\{-dt_{c1}^{HD}\right\} + \exp\left\{-dt_{c1}^{HD}\right\}\frac{(1-\beta)L\left(1-\beta-2\alpha(1-\alpha)\right)}{\alpha(1-\alpha)-\beta(1-\beta)} > 0,$$

$$\because \beta < \alpha < \frac{1}{2} \Longrightarrow \alpha(1-\alpha) > \beta(1-\beta), \quad 1-\beta-2\alpha(1-\alpha) > 0,$$

$$\frac{\partial E[N]}{\partial d} = \left(\beta(1-\beta)-\alpha(1-\alpha)\right)\frac{-t_{c1}^{HD}\exp\left\{-dt_{c1}^{HD}\right\}d - \exp\left\{-dt_{c1}^{HD}\right\}}{d^2}$$

$$+\frac{-\alpha(1-\alpha)}{d^2} = \frac{\alpha(1-\alpha)}{d^2}\left(-1+\exp\left\{-dt_{c1}^{HD}\right\}\left(dt_{c1}^{HD}+1\right)\right)$$

$$-\beta(1-\beta)\frac{\exp\left\{-dt_{c1}^{HD}\right\}\left(dt_{c1}^{HD}+1\right)}{d^2}$$

$$= \frac{\alpha(1-\alpha)}{d^2}\exp\left\{-dt_{c1}^{HD}\right\}\left(-\exp\left\{dt_{c1}^{HD}\right\} + \left(1+dt_{c1}^{HD}\right)\right)$$

$$-\beta(1-\beta)\frac{\exp\left\{-dt_{c1}^{HD}\right\}\left(dt_{c1}^{HD}+1\right)}{d^2}$$

$$< \frac{\alpha(1-\alpha)}{d^2}\exp\left\{-dt_{c1}^{HD}\right\}\left(-\left(1+dt_{c1}^{HD}\right) + \left(1+dt_{c1}^{HD}\right)\right)$$

$$-\beta(1-\beta)\frac{\exp\left\{-dt_{c1}^{HD}\right\}\left(dt_{c1}^{HD}+1\right)}{d^2}$$

$$= -\beta(1-\beta)\frac{\exp\left\{-dt_{c1}^{HD}\right\}\left(dt_{c1}^{HD}+1\right)}{d^2} < 0,$$

$$\frac{\partial E[N]}{\partial L} = \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d}e^{-dt_{c1}^{HD}}\frac{-d(1-\beta)^2}{\alpha(1-\alpha)-\beta(1-\beta)} = e^{-dt_{c1}^{HD}}(1-\beta)^2 > 0.$$

For the red term, we used a mathematical conclusion: $e^x > 1+x$ for any $x > 0$. Therefore, $E[N]$ is an increasing function of α, β and L, a decreasing function of d. From mathematical point of view, this can be explained as follows. Recall that the protein

production rate changes from $\alpha(1-\alpha)$ to $\beta(1-\beta)$ at t_{c1}^{HD}, and t_{c1}^{HD} is proportional to L. The increase of α, β, L is equivalent to an increase of the average protein production rate. Intuitively, the increase of d means a decrease of an average mRNA lifetime, thus a decrease of an average protein production. $E[N]$ tends to $\dfrac{\alpha(1-\alpha)}{d}$ with the increasing of β or L:

$$\lim_{\beta\to\alpha} E[N] = \frac{\alpha(1-\alpha)}{d} + \frac{\alpha(1-\alpha)-\alpha(1-\alpha)}{d}\lim_{\beta\to\alpha}\exp\left\{-dt_{c1}^{HD}\right\} = \frac{\alpha(1-\alpha)}{d},$$
$$\lim_{L\to\infty} E[N] = \frac{\alpha(1-\alpha)}{d} + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d}\lim_{L\to\infty}\exp\left\{-dt_{c1}^{HD}\right\} = \frac{\alpha(1-\alpha)}{d},$$

since as $\beta\to\alpha$ or $L\to\infty$, $t_{c1}^{HD}\to\infty$, the average protein production rate is given by $\alpha(1-\alpha)$. Moreover, it is not difficult to show that:

$$E[N] = \frac{\alpha(1-\alpha)}{d} + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d}\exp\left\{-dt_{c1}^{HD}\right\} < \frac{\alpha(1-\alpha)}{d},$$

since $\beta(1-\beta)-\alpha(1-\alpha)<0$ when $1/2>\alpha>\beta$ and

$$\begin{aligned} E[N] =& \frac{\alpha(1-\alpha)}{d} + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d}\exp\left\{-dt_{c1}^{HD}\right\} \\ >& \frac{\alpha(1-\alpha)}{d} + \frac{\beta(1-\beta)-\alpha(1-\alpha)}{d} = \frac{\beta(1-\beta)}{d}. \end{aligned}$$

Thus, $E[N]$ is bounded by:

$$\frac{\beta(1-\beta)}{d} < E[N] < \frac{\alpha(1-\alpha)}{d}, \tag{3.3.5}$$

which is easy to understand because that the average protein production rate is bounded by $\alpha(1-\alpha)$ and $\beta(1-\beta)$. Now we consider the peak of the probability density function (3.3.3) noted above. The critical point at which this peak occurs is given by:

$$n = \alpha(1-\alpha)t_{c1}^{HD} = \frac{\alpha(1-\alpha)(1-\beta)^2 L}{\alpha(1-\alpha)-\beta(1-\beta)} = \frac{(1-\beta)^2 L}{1-\beta(1-\beta)/\left(\alpha(1-\alpha)\right)}, \tag{3.3.6}$$

which is obviously a decreasing function of α and an increasing function of β and L: smaller α, larger β, L give larger t_{c1}^{HD}, which means later the critical point occurs.

The value of f_N at the critical point is $H = f_N\left(\alpha(1-\alpha)t_{c1}^{HD}\right)$, i.e.,

$$\begin{aligned} H &= \frac{d}{\beta(1-\beta)}\exp\left\{-d\left(\frac{\alpha(1-\alpha)t_{c1}^{HD}}{\beta(1-\beta)} - \frac{1-\beta}{\beta}L\right)\right\} \\ &= \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\} = \frac{d}{\beta(1-\beta)}\exp\left\{-dt_{c1}^{HD}\right\}. \end{aligned} \tag{3.3.7}$$

Taking derivatives of H with respect to α, β, d, L, respectively (note $1/2 > \alpha > \beta$) yields:

$$\begin{aligned} \frac{\partial H}{\partial \alpha} &= \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}\frac{d(1-\beta)^2L(1-2\alpha)}{\left(\alpha(1-\alpha)-\beta(1-\beta)\right)^2} > 0, \\ \frac{\partial H}{\partial \beta} &= \frac{-d(1-2\beta)}{\left(\beta(1-\beta)\right)^2}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\} \\ &\quad + \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}\frac{d(1-\beta)L\left(2\alpha(1-\alpha)+\beta-1\right)}{\alpha(1-\alpha)-\beta(1-\beta)} < 0, \\ \frac{\partial H}{\partial d} &= \frac{1}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\} \\ &\quad + \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}\frac{-(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)} \\ &= \frac{\exp\left\{\dfrac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}}{\alpha(1-\alpha)-\beta(1-\beta)}\left(\alpha(1-\alpha)-\beta(1-\beta)-d(1-\beta)^2L\right) \\ &> 0 \quad \text{if} \quad d < \frac{\alpha(1-\alpha)-\beta(1-\beta)}{(1-\beta)^2L}, \\ &< 0 \quad \text{if} \quad d > \frac{\alpha(1-\alpha)-\beta(1-\beta)}{(1-\beta)^2L}, \\ \frac{\partial H}{\partial L} &= \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(1-\beta)^2L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}\frac{-d(1-\beta)^2}{\alpha(1-\alpha)-\beta(1-\beta)} < 0. \end{aligned}$$

Therefore, H is increasing with α, decreasing with β and L, this is because within a

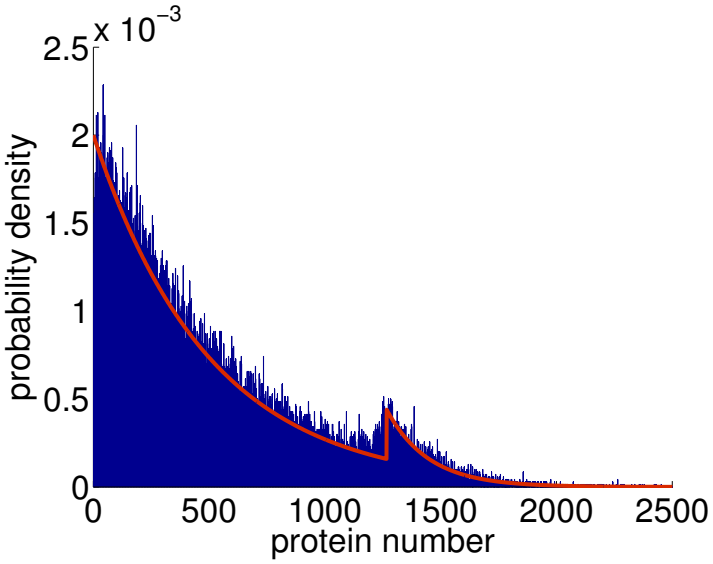


Figure 3.7: **Probability density distribution of protein for high density phase under tracking degradation:** $\alpha \geqslant 1/2$. Red curve is probability density function (3.3.8). Blue bars are probability density distribution from simulations (please refer to the text for details). $\alpha = 0.8, \beta = 0.1, L = 1000, d = 0.0005$.

fixed mRNA lifetime span (d), more mRNAs are able to produce the protein number at the critical point value with smaller t_{c1}^{HD} (larger α, smaller β, L). H is increasing with d if $d < \dfrac{\alpha(1-\alpha)-\beta(1-\beta)}{(1-\beta)^2 L}$, and decreasing if $d > \dfrac{\alpha(1-\alpha)-\beta(1-\beta)}{(1-\beta)^2 L}$. This is non-intuitive. With the increase of d, the average lifetime of an mRNA decreases. Consequently, the average mRNA lifetime gets close to t_{c1}^{HD} first and then moves away from t_{c1}^{HD} due to the further decreasing of d. Therefore, with d increases, more mRNAs produce proteins at the critical point value first and then less mRNAs produce proteins at the critical point value . Note that H is decreasing to 0 with the increasing of β, L and d:

$$\lim_{\beta\to\alpha \text{ or } L\to\infty} H = \lim_{\beta\to\alpha \text{ or } L\to\infty} \frac{d}{\beta(1-\beta)} \lim_{\beta\to\alpha \text{ or } L\to\infty} \exp\left\{-dt_{c1}^{HD}\right\} = 0,$$

since as $\beta \to \alpha$ or $L \to \infty$, $t_{c1}^{HD} \to \infty$.

Moreover,

$$\begin{aligned}\lim_{d\to\infty} H &= \lim_{d\to\infty} \frac{1}{\beta(1-\beta)} \frac{d}{\exp\left\{\dfrac{d(1-\beta)^2 L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}} \\ &= \lim_{d\to\infty} \frac{1}{\beta(1-\beta)} \frac{1}{\exp\left\{\dfrac{d(1-\beta)^2 L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\} \dfrac{(1-\beta)^2 L}{\alpha(1-\alpha)-\beta(1-\beta)}} \\ &= \frac{1}{\dfrac{\beta(1-\beta)^3 L}{\alpha(1-\alpha)-\beta(1-\beta)} \lim_{d\to\infty} \exp\left\{\dfrac{d(1-\beta)^2 L}{\alpha(1-\alpha)-\beta(1-\beta)}\right\}} = 0.\end{aligned}$$

Thus, one can mediate where the peak occurs and the value of f_N at the critical point by carefully choosing the parameters to control the likelihood of protein production sitting around the critical point. For example, in Figure 3.6B, the likelihood around 560 - 600 increases sharply.

3.3.4 High Density Phase: $\alpha \geqslant 1/2$

Analysis and properties of the high density phase when $\alpha \geqslant 1/2$ is similar to those of the high density phase when $\alpha < 1/2$, so we may omit some insignificant details. The interpretation of the results is also omitted, please refer back to case of $\alpha < 1/2$ for a better understanding.

Recall that when $\alpha \geqslant \dfrac{1}{2}$, N is approximately given by (see (2.3.6)):

$$N = \begin{cases} \dfrac{1}{4}T, & T < t_{c2}^{HD}, \\ \beta(1-\beta)T + \dfrac{3-4\beta}{4}L, & T \geqslant t_{c2}^{HD}, \end{cases} \qquad t_{c2}^{HD} = \frac{(3-4\beta)L}{1-4\beta(1-\beta)}.$$

The probability density function for protein number $N = n$ can be derived using Lemma

3.1:

$$f_N(n) = \begin{cases} 4d\exp\{-4dn\}, & n < \frac{1}{4}t_{c2}^{HD}, \\ \frac{d}{\beta(1-\beta)}\exp\left\{-d\left(\frac{n}{\beta(1-\beta)} - \frac{(3-4\beta)L}{4\beta(1-\beta)}\right)\right\}, & n \geqslant \frac{1}{4}t_{c2}^{HD}. \end{cases} \tag{3.3.8}$$

From Figure 3.7, we see that the probability density function (3.3.8) has two branches. One branch is an exponential distribution with parameter $4d$, which can be regarded as a distribution of protein being produced at rate $\frac{1}{4}$. The other branch can be obtained by shifting an exponential distribution with parameter $\frac{d}{\beta(1-\beta)}$, which can be regarded as a distribution of protein being produced at rate $\beta(1-\beta)$. Again, we observed a critical point at which there occurs a peak. This is because at the critical point, the protein production rate changes from $\frac{1}{4}$ to $\beta(1-\beta)$ (as we discussed in Section 2.3.2.2). We will discuss the properties of the critical point later.

Let $E[N]$ to be the expected value of (3.3.8), then $E[N]$ is given by:

$$E[N] = \int_0^\infty n f_N(n) dn.$$

We obtain:

$$\begin{aligned} E[N] &= \int_0^\infty n f_N(n) dn = \int_0^{t_{c2}^{HD}/4} n4d\exp\{-4dn\} dn \\ &\quad + \int_{t_{c2}^{HD}/4}^\infty n\frac{d}{\beta(1-\beta)}\exp\left\{-d\left(\frac{n}{\beta(1-\beta)} - \frac{(3-4\beta)L}{4\beta(1-\beta)}\right)\right\} dn \\ &= \frac{1}{4d} - \frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\}. \end{aligned} \tag{3.3.9}$$

Taking derivatives of $E(n)$ as defined in (3.3.9) with respect to α, β, d, L, respectively yields:

$$\frac{\partial E[N]}{\partial \alpha} = 0,$$

$$\begin{aligned}
\frac{\partial E[N]}{\partial \beta} &= \frac{4-8\beta}{4d}\exp\left\{-dt_{c2}^{HD}\right\} \\
&\quad - \frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\}\frac{\partial}{\partial \beta}\left(\frac{-d(3-4\beta)L}{1-4\beta(1-\beta)}\right) \\
&= \frac{(2\beta-1)^2+2dL(1-\beta)}{(1-2\beta)d}\exp\left\{-dt_{c2}^{HD}\right\} > 0, \\
\frac{\partial E[N]}{\partial d} &= \frac{-1}{4d^2} + \frac{1-4\beta(1-\beta)}{4d^2}\exp\left\{-dt_{c2}^{HD}\right\} + \frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\}t_{c2}^{HD} \\
&= \frac{-1}{4d^2} + \frac{1}{4d^2}\exp\left\{-dt_{c2}^{HD}\right\}(1-2\beta)^2\left(1+dt_{c2}^{HD}\right) \\
&< \frac{-1}{4d^2} + \frac{1}{4d^2}\exp\left\{-dt_{c2}^{HD}\right\}\exp\left\{dt_{c2}^{HD}\right\} = \frac{-1}{4d^2} + \frac{1}{4d^2} < 0, \\
&\because (1-2\beta)^2 < 1, \\
\frac{\partial E[N]}{\partial L} &= -\frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\}\frac{-d(3-4\beta)}{1-4\beta(1-\beta)} \\
&= \frac{3-4\beta}{4}\exp\left\{-dt_{c2}^{HD}\right\} > 0.
\end{aligned}$$

Therefore, $E[N]$ is independent of α. $E[N]$ is an increasing function of β and L, is a decreasing function of d. It is easy to show that:

$$\begin{aligned}
\lim_{\beta\to 1/2} E[N] &= \frac{1}{4d} - \frac{1-4\frac{1}{4}}{4d}\exp\left\{-dt_{c2}^{HD}\right\} = \frac{1}{4d}, \\
\lim_{L\to\infty} E[N] &= \frac{1}{4d} - \frac{1-4\beta(1-\beta)}{4d}\lim_{L\to\infty}\exp\left\{-dt_{c2}^{HD}\right\} = \frac{1}{4d}, \\
\lim_{d\to\infty} E[N] &= \frac{1}{\infty} - \frac{1-4\beta(1-\beta)}{\infty}\lim_{d\to\infty}\exp\left\{-dt_{c2}^{HD}\right\} = 0.
\end{aligned}$$

Moreover, it is not difficult to show that:

$$E[N] = \frac{1}{4d} - \frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\} < \frac{1}{4d},$$

since $1-4\beta(1-\beta) > 0$ when $\beta < 1/2$, and

$$E[N] = \frac{1}{4d} - \frac{1-4\beta(1-\beta)}{4d}\exp\left\{-dt_{c2}^{HD}\right\} > \frac{1}{4d} - \frac{1-4\beta(1-\beta)}{4d} = \frac{\beta(1-\beta)}{d}.$$

Thus, $E[N]$ is bounded by:

$$\frac{\beta(1-\beta)}{d} < E[N] < \frac{1}{4d}. \tag{3.3.10}$$

Now we consider the peak of the probability density function (3.3.8) noted above. The critical point at which this peak occurs is given by:

$$n = \frac{1}{4}t_{c2}^{HD} = \frac{(3-4\beta)L}{4(1-2\beta)^2}, \tag{3.3.11}$$

which obviously is linearly increasing with L. Besides:

$$\frac{\partial n}{\partial \beta} = \frac{2L(1-\beta)}{(1-2\beta)^3} > 0.$$

Thus, the critical point is increasing with β.

The value of f_N at the critical point is $H = f_N\left(\frac{1}{4}t_{c2}^{HD}\right)$, i.e.,

$$\begin{aligned} H &= \frac{d}{\beta(1-\beta)}\exp\left\{-d\left(\frac{\frac{1}{4}t_{c2}^{HD}}{\beta(1-\beta)} - \frac{(3-4\beta)L}{4\beta(1-\beta)}\right)\right\} \\ &= \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\} = \frac{d}{\beta(1-\beta)}\exp\left\{-dt_{c2}^{HD}\right\}. \end{aligned} \tag{3.3.12}$$

Taking derivatives of H with respect to α, β, L, d, respectively (note $\alpha \geqslant 1/2 > \beta$) yields:

$$\begin{aligned} &\frac{\partial H}{\partial \alpha} = 0, \\ &\frac{\partial H}{\partial \beta} = \frac{-d(1-2\beta)}{\beta^2(1-\beta)^2}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\} \\ &\quad + \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\}\frac{-8dL(1-\beta)}{(1-2\beta)^3} \end{aligned}$$

$$= \frac{-d(1-2\beta)}{\beta^2(1-\beta)^2}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\} - \frac{8d^2L}{\beta(1-2\beta)^3}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\} < 0,$$

$$\frac{\partial H}{\partial L} = \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\}\frac{-d(3-4\beta)}{(1-2\beta)^2} < 0,$$

$$\frac{\partial H}{\partial d} = \frac{1}{\beta(1-\beta)}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\} + \frac{d}{\beta(1-\beta)}\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\}\frac{-(3-4\beta)L}{(1-2\beta)^2}$$

$$= \frac{\exp\left\{\frac{-d(3-4\beta)L}{(1-2\beta)^2}\right\}}{\beta(1-\beta)(1-2\beta)^2}\Big((1-2\beta)^2 - (3-4\beta)dL\Big)$$

$$> 0 \quad \text{if} \quad d < \frac{(1-2\beta)^2}{(3-4\beta)L}, \qquad < 0 \quad \text{if} \quad d > \frac{(1-2\beta)^2}{(3-4\beta)L}.$$

Therefore, H is decreasing with β and L, first increasing then decreasing with d. Moreover, with the increasing of β, L, d, H tends to 0:

$$\lim_{\beta\to 1/2} H = \frac{d}{1/4}\lim_{\beta\to 1/2}\exp\left\{-dt_{c2}^{HD}\right\} = 0, \lim_{L\to\infty} H = \frac{d}{\beta(1-\beta)}\lim_{L\to\infty}\exp\left\{-dt_{c2}^{HD}\right\} = 0,$$

$$\lim_{d\to\infty} H = \lim_{d\to\infty}\frac{1}{\beta(1-\beta)}\frac{d}{\exp\left\{\frac{d(3-4\beta)L}{(1-2\beta)^2}\right\}}$$

$$= \frac{1}{\beta(1-\beta)}\frac{1}{\frac{(3-4\beta)L}{(1-2\beta)^2}}\lim_{d\to\infty}\frac{1}{\exp\left\{\frac{d(3-4\beta)L}{(1-2\beta)^2}\right\}} = 0.$$

Thus, one can mediate where the peak occurs and the value of f_N at the critical point by carefully choosing the parameters to control the likelihood of protein production sitting around the critical point.

3.4 Protein Distribution under Rapid Degradation

As we introduced, this type of mRNA degradation is rapid after the initial degradosome binding event. For rapid degradation, at the time of degradosome binding, we assume that the mRNA is degraded immediately. Therefore, the number of proteins produced from one mRNA is equivalent to number of ribosomes that leave the mRNA before

degradosome binding occurs. Recall that in Section 2.3, we approximated the number of ribosome particle leaving the lattice by N_2, where N_2 is a function of time, t. In fact, by replacing t with T (T is the random variable representing degradosome binding time), N_2 can be used to approximate the expected number of proteins per mRNA, N.

3.4.1 Low Density Phase

Recall from (2.3.2) that in the low density phase, N is approximately given by:

$$N = \begin{cases} 0, & 0 < T \leqslant L, \\ \dfrac{\alpha(1-\alpha)}{2}(T-L), & L < T < t_c^{LD}, \qquad t_c^{LD} = \dfrac{1+\alpha}{1-\alpha}L. \\ \alpha(1-\alpha)T - \alpha L, & T \geqslant t_c^{LD}, \end{cases}$$

The probability density function of T is given by (3.2.1). We apply Lemma 3.1 to obtain the probability density function for protein number N directly. For $L < T < t_c^{LD}$, namely, $N = n \in \left(\dfrac{\alpha(1-\alpha)}{2}(L-L), \dfrac{\alpha(1-\alpha)}{2}(t_c^{LD}-L)\right) = \left(0, \alpha^2 L\right)$, the corresponding probability density function is:

$$\begin{aligned} f_N(n) &= \frac{1}{\dfrac{\alpha(1-\alpha)}{2}} d \exp\left\{-d\left(\frac{n-\left(-\dfrac{\alpha(1-\alpha)}{2}L\right)}{\dfrac{\alpha(1-\alpha)}{2}}\right)\right\} \\ &= \frac{2d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{2n}{\alpha(1-\alpha)} + L\right)\right\}. \end{aligned}$$

For $T \geqslant t_c^{LD}$, namely, $N = n \geqslant \alpha^2 L$:

$$f_N(n) = \frac{d \exp\left\{-d\left(\frac{n-(-\alpha L)}{\alpha(1-\alpha)}\right)\right\}}{\alpha(1-\alpha)} = \frac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{n+\alpha L}{\alpha(1-\alpha)}\right)\right\}.$$

For $0 < T \leqslant L$, $N = 0 \cdot T + 0 = 0$, thus we cannot obtain the corresponding probability

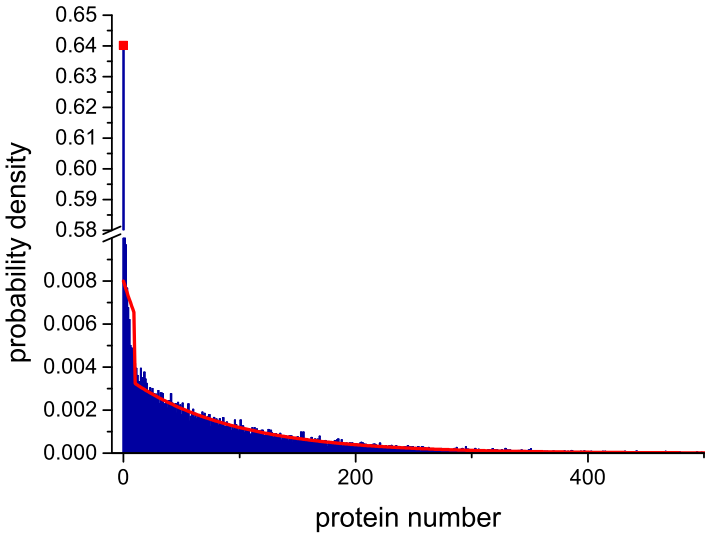


Figure 3.8: **Probability density distribution of protein for low density phase: rapid degradation.** Red curve is probability density function (3.4.2). Red dot is the probability of producing no protein (3.4.3). Blue bars are probability density distribution from simulations (please refer to the text for details). $\alpha = 0.1, \beta = 0.8, L = 1000, d = 0.001$.

density function by applying Lemma 3.1. In practice, the probability that the mRNA is degraded without producing any proteins, $P(0)$, is certainly not zero. However, in practice and in simulation, the protein number is discrete while in our model, N is continuous. In this sense, we define:

$$P(0) := P(0 \leqslant n < 1) = \int_0^1 f_N(n)dn = 1 - \int_1^\infty f_N(n)dn. \tag{3.4.1}$$

Therefore, if $\alpha^2 L > 1$,

$$\begin{aligned} P(0) &= 1 - \int_1^\infty f_N(n)dn = 1 - \int_1^{\alpha^2 L} f_N(n)dn - \int_{\alpha^2 L}^\infty f_N(n)dn \\ &= 1 - \int_1^{\alpha^2 L} \frac{2d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{2n}{\alpha(1-\alpha)} + L\right)\right\} dn \\ &\quad - \int_{\alpha^2 L}^\infty \frac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{n+\alpha L}{\alpha(1-\alpha)}\right)\right\} dn \end{aligned}$$

$$= 1 - \exp\left\{-d\left(L + \frac{2}{\alpha(1-\alpha)}\right)\right\}.$$

If $\alpha^2 L \leqslant 1$,

$$\begin{aligned} P(0) &= 1 - \int_1^\infty f_N(n)dn = 1 - \int_1^\infty \frac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{n+\alpha L}{\alpha(1-\alpha)}\right)\right\} dn \\ &= 1 - \exp\left\{-d\left(\frac{1+\alpha L}{\alpha(1-\alpha)}\right)\right\}. \end{aligned}$$

To summarise, the probability density function for protein number is as follows,

$$f_N(n) = \begin{cases} \dfrac{2d}{\alpha(1-\alpha)} \exp\left\{-d\left(\dfrac{2n}{\alpha(1-\alpha)} + L\right)\right\}, & 0 < n < \alpha^2 L, \\ \dfrac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\dfrac{n+\alpha L}{\alpha(1-\alpha)}\right)\right\}, & n > \alpha^2 L, \end{cases} \tag{3.4.2}$$

and

$$P(0) = \begin{cases} 1 - \exp\left\{-d\left(L + \dfrac{2}{\alpha(1-\alpha)}\right)\right\} & n = 0, \quad \text{if } \alpha^2 L > 1, \\ 1 - \exp\left\{-d\left(\dfrac{1+\alpha L}{\alpha(1-\alpha)}\right)\right\} & n = 0, \quad \text{if } \alpha^2 L < 1. \end{cases} \tag{3.4.3}$$

From Figure 3.8, we see that the simulation and the approximated probability density function are in good agreement. Note that the dot at $n = 0$ represents the probability to produce no protein. We also observed that the probability density function is discontinuous at the critical point $n = \alpha^2 L$. We will study these properties later.

Let $E[N]$ denote the expected value of (3.4.2), then $E[N]$ is given by:

$$E[N] = \int_0^\infty n f_N(n)dn$$

$$= \int_0^{\alpha^2 L} n \frac{2d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{2n}{\alpha(1-\alpha)} + L\right)\right\} dn$$
$$+ \int_{\alpha^2 L}^{\infty} n \frac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{n+\alpha L}{\alpha(1-\alpha)}\right)\right\} dn$$
$$= \frac{\alpha(1-\alpha)}{2d}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\} + \exp\{-dL\}\right). \tag{3.4.4}$$

Next, we study the properties of the expected value $E[N]$. Taking derivatives of $E[N]$ as defined in (3.4.4) with respect to α, β, d, L, respectively yields:

$$\frac{\partial E[N]}{\partial \alpha} = \frac{1-2\alpha}{2d}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\} + \exp\{-dL\}\right)$$
$$+ \frac{\alpha(1-\alpha)}{2d}\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}\left(-dL\frac{1-\alpha+1+\alpha}{(1-\alpha)^2}\right)$$
$$= \frac{1-2\alpha}{2d}\exp\left\{-\frac{(1+\alpha)dL}{1-\alpha}\right\}\left(\exp\left\{\frac{2\alpha dL}{1-\alpha}\right\} - \left(1 + \frac{2\alpha dL}{(1-\alpha)(1-2\alpha)}\right)\right)$$
$$> 0 \quad \text{if} \quad \exp\left\{\frac{2\alpha dL}{1-\alpha}\right\} > \left(1 + \frac{2\alpha dL}{(1-\alpha)(1-2\alpha)}\right),$$
$$< 0 \quad \text{if} \quad \exp\left\{\frac{2\alpha dL}{1-\alpha}\right\} < \left(1 + \frac{2\alpha dL}{(1-\alpha)(1-2\alpha)}\right),$$
$$\frac{\partial E[N]}{\partial \beta} = 0,$$
$$\frac{\partial E[N]}{\partial d} = \frac{-\alpha(1-\alpha)}{2d^2}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\} + \exp\{-dL\}\right)$$
$$- \frac{\alpha(1-\alpha)}{2d}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}\frac{(1+\alpha)L}{1-\alpha} + \exp\{-dL\}L\right) < 0,$$
$$\frac{\partial E[N]}{\partial L} = -\frac{\alpha(1-\alpha)}{2d}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}\frac{-(1+\alpha)d}{1-\alpha} + \exp\{-dL\}d\right) < 0.$$

Therefore, if $\exp\left\{\frac{2\alpha dL}{1-\alpha}\right\} > \left(1 + \frac{2\alpha dL}{(1-\alpha)(1-2\alpha)}\right)$, $E[N]$ is increasing with α, otherwise is decreasing with α. This is non-intuitive. Biological interpretation remains unaccounted for. $E[N]$ is decreasing with d and L. This is because with the increase of d, the average lifetime of an mRNA decreases; with the increase of L, it takes longer

time for a ribosome to complete its translation. Moreover, $E[N]$ tends to 0 with the increasing of d, L:

$$\lim_{L\to\infty} E[N] = \frac{\alpha(1-\alpha)}{2d}\left(\lim_{L\to\infty}\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}+\lim_{L\to\infty}\exp\{-dL\}\right)=0,$$
$$\lim_{d\to\infty} E[N] = \frac{\alpha(1-\alpha)}{\infty}\left(\lim_{d\to\infty}\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}+\lim_{d\to\infty}\exp\{-dL\}\right)=0.$$

Moreover, it is easy to show that $E[N]$ is bounded by $\dfrac{\alpha(1-\alpha)}{d}$ since:

$$\begin{aligned} E[N] &= \frac{\alpha(1-\alpha)}{2d}\left(\exp\left\{\frac{-dL(1+\alpha)}{1-\alpha}\right\}+\exp\{-dL\}\right) \\ &< \frac{\alpha(1-\alpha)}{2d}(1+1) = \frac{\alpha(1-\alpha)}{d}, \end{aligned} \tag{3.4.5}$$

since the average protein production rate is bounded by $\alpha(1-\alpha)$.

Let $Z = P(0)$, from (3.4.3), even without taking derivatives of Z, we are able to determine that Z is decreasing with α and increasing with d and L. This is because with smaller α and larger d, L, an mRNA is more likely to produce no protein before the mRNA is degraded.

Moreover, Z tends to a limiting level (suppose $L \gg 4$ such that $\alpha^2 L \gg 1$ when $\alpha \to 1/2$):

$$\lim_{\alpha\to 1/2} Z = \lim_{\alpha\to 1/2}\left(1-\exp\left\{-d\left(L+\frac{2}{\alpha(1-\alpha)}\right)\right\}\right) = 1-e^{-d(L+8)},$$

and Z tends to 1 with the increasing of d and L:

$$\lim_{d,L\to\infty} Z = 1-\lim_{d,L\to\infty}\exp\left\{-d\left(L+\frac{2}{\alpha(1-\alpha)}\right)\right\} = 1.$$

The probability density function (3.4.2) is discontinuous at a critical point given by $\alpha^2 L$. Let $C = \alpha^2 L$, and $H = f_N\left(C^-\right)$. The critical point $C = \alpha^2 L$ is clearly positively correlated with α and L. $H = f_N\left(C^-\right)$ is given by:

$$\begin{aligned} H &= \frac{2d}{\alpha(1-\alpha)} \exp\left\{ -d\left(\frac{2\alpha^2 L}{\alpha(1-\alpha)} + L \right) \right\} \\ &= \frac{2d}{\alpha(1-\alpha)} \exp\left\{ \frac{-(1+\alpha)dL}{1-\alpha} \right\}. \end{aligned} \tag{3.4.6}$$

Clearly H is decreasing with α and L. This is because with the increase of α and L, the critical point value increases, less mRNAs produce that amount of proteins during their lifetimes. The corresponding limitations of H is therefore given by:

$$\begin{aligned} \lim_{\alpha \to 1/2} H &= \frac{2d \exp\left\{ \frac{-(1+1/2)dL}{1-1/2} \right\}}{1/2(1-1/2)} = 8de^{-3dL}, \\ \lim_{L \to \infty} H &= \frac{2d}{\alpha(1-\alpha)} \lim_{L \to \infty} \exp\left\{ \frac{-(1+\alpha)dL}{1-\alpha} \right\} = 0. \end{aligned}$$

Take derivative of H with respect to d:

$$\begin{aligned} \frac{\partial H}{\partial d} &= \frac{2\exp\left\{ \frac{-(1+\alpha)dL}{1-\alpha} \right\}}{\alpha(1-\alpha)} - \frac{2d}{\alpha(1-\alpha)} \exp\left\{ \frac{-(1+\alpha)dL}{1-\alpha} \right\} \frac{(1+\alpha)L}{1-\alpha} \\ &= \frac{2\exp\left\{ \frac{-(1+\alpha)dL}{1-\alpha} \right\}}{\alpha(1-\alpha)} \left(1 - \frac{1+\alpha}{1-\alpha} dL \right) \\ &> 0 \quad \text{if} \quad d < \frac{1-\alpha}{(1+\alpha)L}, \qquad\qquad < 0 \quad \text{if} \quad d > \frac{1-\alpha}{(1+\alpha)L}. \end{aligned}$$

Hence, H is first increasing then decreasing with d. Note that $\dfrac{1-\alpha}{(1+\alpha)L}$ is the reciprocal of t_c^{LD} and the average lifetime of an mRNA is $\dfrac{1}{d}$. Namely, if the average lifetime of an mRNA is shorter than t_c^{LD}, with the increase of d, less mRNAs produce proteins

at the critical value and vise versa. H tends to 0 with increasing d:

$$\begin{aligned}\lim_{d\to\infty} H &= \lim_{d\to\infty} \frac{2}{\alpha(1-\alpha)} \frac{d}{\exp\left\{\dfrac{(1+\alpha)dL}{1-\alpha}\right\}} \\ &= \frac{2}{\alpha(1-\alpha)} \frac{1}{\dfrac{(1+\alpha)L}{1-\alpha}} \lim_{d\to\infty} \frac{1}{\exp\left\{\dfrac{(1+\alpha)dL}{1-\alpha}\right\}} = 0.\end{aligned}$$

Now we consider $f_N(C^+)$, which is given by:

$$f_N(C^+) = \frac{d}{\alpha(1-\alpha)} \exp\left\{-d\left(\frac{\alpha^2 L+\alpha L}{\alpha(1-\alpha)}\right)\right\} = \frac{d\exp\left\{\dfrac{-(1+\alpha)dL}{1-\alpha}\right\}}{\alpha(1-\alpha)} = \frac{H}{2}.$$

Hence, as we can see, the probability density function (3.4.2) is discontinuous at the critical point $C = \alpha^2 L$ and the probability density is halved as N increases through this value.

3.4.2 High Density Phase

Recall from (2.3.4) and (2.3.6) that in the high density phase, N is approximated by:

$$N = \begin{cases} 0, & T \leqslant L, \\ \beta(1-\beta)(T-L), & T > L. \end{cases}$$

The probability density function of T is given by (3.2.1). We apply Lemma 3.1 to obtain the probability density function for protein number N directly. For $T > L$, $N = n > \beta(1-\beta)(L-L) = 0$, the corresponding probability density function is:

$$f_N(n) = \frac{d}{\beta(1-\beta)} \exp\left\{-d\left(\frac{n}{\beta(1-\beta)} + L\right)\right\}.$$

For $0 < T \leqslant L$, $N = 0 = 0 \cdot T + 0$, thus the corresponding probability density function cannot be obtained by appling Lemma 3.1. The probability of producing no protein is:

$$\begin{aligned} P(0) &= 1 - \int_1^\infty f_N(n)dn = 1 - \int_1^\infty \frac{d}{\beta(1-\beta)} \exp\left\{-d\left(\frac{n}{\beta(1-\beta)} + L\right)\right\} dn \\ &= 1 - \exp\left\{-d\left(\frac{1}{\beta(1-\beta)} + L\right)\right\}. \end{aligned} \tag{3.4.7}$$

To summarise,

$$f_N(n) = \frac{d}{\beta(1-\beta)} \exp\left\{-d\left(\frac{n}{\beta(1-\beta)} + L\right)\right\}, \qquad n \in \mathbb{R}^+, \tag{3.4.8}$$

$$P(0) = 1 - \exp\left\{-d\left(\frac{1}{\beta(1-\beta)} + L\right)\right\}. \tag{3.4.9}$$

We run 50000 simulations using the Monte Carlo method introduced in Section 3.1 and 2.4.2. The result is shown in Figure 3.9. The mean value of simulation is 105.2275, whilst the expected value of (3.4.8) is 109.1775 (whose expression will be given in (3.4.10)). Thus our simulation and the approximated probability density function appear to match well.

Let $E[N]$ denote expected value of (3.4.8), then $E[N]$ is given by:

$$\begin{aligned} E[N] &= \int_0^\infty n f_N(n)dn = \int_0^\infty n \frac{d}{\beta(1-\beta)} \exp\left\{-d\left(\frac{n}{\beta(1-\beta)} + L\right)\right\} dn \\ &= \frac{\beta(1-\beta)}{d} \exp\{-dL\}. \end{aligned} \tag{3.4.10}$$

$E[N]$ as defined in (3.4.10) is a simple expression, without taking derivatives, we can determine that $E[N]$ is a decreasing function of d, L and an increasing function of β for $\beta < \frac{1}{2}$. This is because with the increase of d, the average lifetime of an mRNA decreases; with the increase of L, a ribosome needs longer time to complete its translation; larger β provides a lager protein production rate. Furthermore, $E[N]$ tends to 0

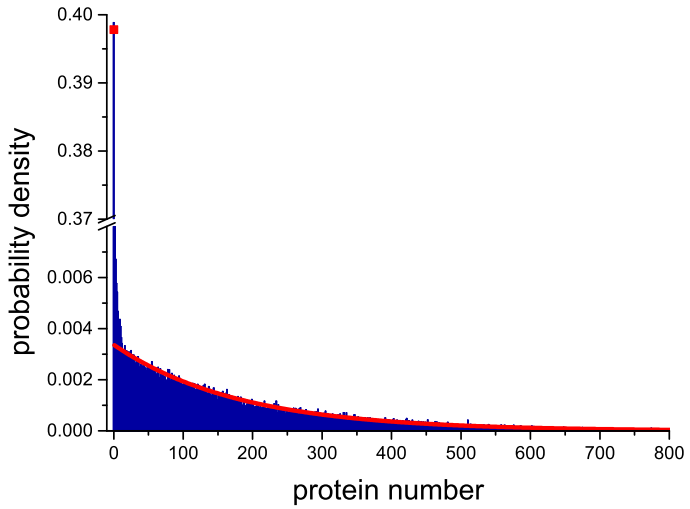


Figure 3.9: **Probability density distribution of protein for high density phase: rapid degradation.** Red curve is probability density function (3.4.8). Red dot is the probability of producing no protein (3.4.9). Blue bars are probability density distribution from simulations. $\alpha = 0.4, \beta = 0.1, L = 1000, d = 0.0005$.

with the increasing of d and L since:

$$\lim_{d\to\infty} E[N] = \frac{\beta(1-\beta)}{\infty} \lim_{d\to\infty} exp\{-dL\} = 0, \lim_{L\to\infty} E[N] = \frac{\beta(1-\beta)}{d} \lim_{L\to\infty} \exp\{-dL\} = 0.$$

It is easy to show that $E[N]$ is bounded by $\beta(1-\beta)/d$:

$$E[N] = \frac{\beta(1-\beta)}{d} \exp\{-dL\} < \frac{\beta(1-\beta)}{d} \qquad (\because \exp\{-dL\} < 1). \tag{3.4.11}$$

Let $Z = P(0)$, which is given by (3.4.9). It is obvious that Z is decreasing with β and increasing with d, L. This is because with smaller β and largerd, L, an mRNA is more likely to produce no protein before the mRNA is degraded. Z tends to following limiting levels:

$$\lim_{\beta\to 1/2} Z = 1 - \exp\left\{-d\left(\frac{1}{1/2(1-1/2)} + L\right)\right\} = 1 - e^{-d(L+4)},$$

$$\lim_{d,L\to\infty} Z = 1 - \lim_{d,L\to\infty} \exp\left\{-d\left(\frac{1}{\beta(1-\beta)}+L\right)\right\} = 1.$$

3.4.3 Maximal Current Phase

Recall from (2.3.10) that in the maximal current phase,

$$N = \begin{cases} 0, & 0 < T \leqslant L, \\ \dfrac{1}{8}(T-L), & L < T \leqslant t_c^{MC}, \\ \dfrac{1}{4}T - \dfrac{1}{2}L, & T > t_c^{MC}, \end{cases} \qquad t_c^{MC} = 3L.$$

The probability density function of T is given by (3.2.1). We apply Lemma 3.1 to obtain the probability density function of protein number N directly. For $L < T \leqslant t_c^{MC}$, $N = n \in \left(\frac{1}{8}(L-L), \frac{1}{8}(3L-L)\right] = \left(0, \frac{L}{4}\right]$, the corresponding probability density function is $f_N(n) = 8d\exp\left\{-d\left(8n+L\right)\right\}$. For $T > t_c^{MC}$, $N = n > \dfrac{L}{4}$, the corresponding probability density function is $f_N(n) = 4d\exp\left\{-d\left(4n+2L\right)\right\}$.

For $0 < T \leqslant L$, $N = 0 \cdot T + 0 = 0$, thus the corresponding probability density function cannot be obtained by applying Lemma 3.1. Probability of producing no protein is:

$$\begin{aligned} P(0) &= 1 - \int_1^\infty f_N(n)dn = 1 - \int_1^{L/4} 8d\exp\left\{-d\left(8n+L\right)\right\}dn \\ &\quad - \int_{L/4}^\infty 4d\exp\left\{-d\left(4n+2L\right)\right\}dn = 1 - \exp\left\{-d(8+L)\right\}. \end{aligned} \tag{3.4.12}$$

To summarise,

$$f_N(n) = \begin{cases} 8d\exp\left\{-d\left(8n+L\right)\right\}, & 0 < n \leqslant L/4, \\ 4d\exp\left\{-d\left(4n+2L\right)\right\}, & n > L/4, \end{cases} \tag{3.4.13}$$

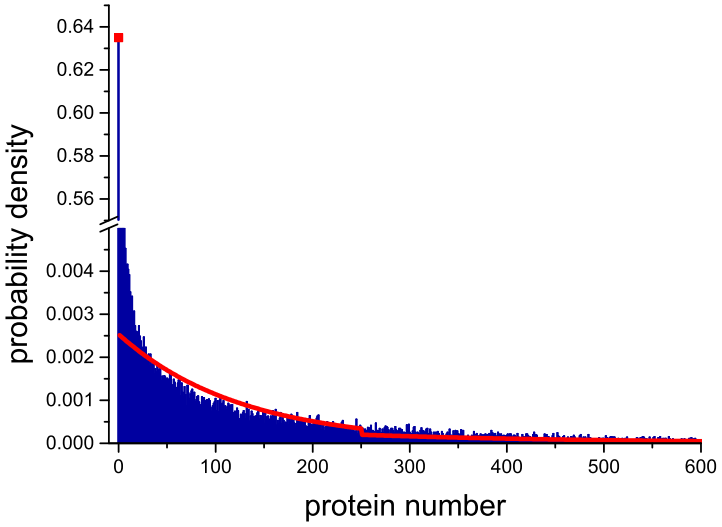


Figure 3.10: **Probability density distribution of protein for maximal current phase: rapid degradation.** Red curve is probability density function (3.4.13). Red dot is the probability of producing no protein (3.4.14). Blue bars are probability density distribution from simulations. $\alpha = 0.6, \beta = 0.6, L = 1000, d = 0.001$.

and

$$P(0) = 1 - \exp\left\{-d(8+L)\right\}. \tag{3.4.14}$$

The simulation result is shown in Figure 3.10. The mean value of simulation is about 54.5178, whilst the expected value of (3.4.13) is 52.2083 (whose expression will be given in (3.4.15)). Though the mean values are close, poor agreement between simulation and approximation is observed. This poor agreement suggests that for the maximal current phase, linear approximation on protein production is not ideal.

Let $E[N]$ denote the expected value of (3.4.13), then $E[N]$ is given by:

$$\begin{aligned} E[N] &= \int_0^\infty n f_N(n) dn \\ &= \int_0^{L/4} n8d \exp\left\{-d\left(8n+L\right)\right\} dn + \int_{L/4}^\infty n4d \exp\left\{-d\left(4n+2L\right)\right\} dn \\ &= \frac{1}{8d}\left(\exp(-dL) + \exp(-3dL)\right). \end{aligned} \tag{3.4.15}$$

It is obvious that $E[N]$ (3.4.15) is decreasing with d and L. With the increase of d, the average lifetime of an mRNA decreases, and with the increase of L, it takes longer time for a ribosome to complete its translation. The corresponding limitations are:

$$\lim_{d\to\infty} E[N] = \frac{1}{\infty}\left(\lim_{d\to\infty}\exp(-dL) + \lim_{d\to\infty}\exp(-3dL)\right) = 0,$$
$$\lim_{L\to\infty} E[N] = \frac{1}{8d}\left(\lim_{L\to\infty}\exp(-dL) + \lim_{L\to\infty}\exp(-3dL)\right) = 0.$$

It is easy to show that $E[N]$ is bounded by $\frac{1}{4d}$:

$$E[N] = \frac{1}{8d}\left(\exp(-dL) + \exp(-3dL)\right) < \frac{1}{8d}(1+1) = \frac{1}{4d}. \tag{3.4.16}$$

Let $Z = P(0)$, which is given by (3.4.14). It is obvious that Z is an increasing function of d and L. This is because that with the increase of $d$and L, more mRNAs produce no proteins before degradation. Z tends to 1:

$$\lim_{d\to\infty} Z = 1 - \lim_{L\to\infty}\exp\left\{-d(8+L)\right\} = 1.$$

The probability density function (3.4.13) is discontinuous at a critical point $n = \frac{L}{4}$. Let $C = \frac{L}{4}$ and $H = f_N\left(C^-\right)$. H is given by:

$$H = 8d\exp\left\{-d\left(8\cdot\frac{L}{4}+L\right)\right\} = 8de^{-3dL}. \tag{3.4.17}$$

H is clearly a decreasing function of L. With the increase of L, the value critical point increases, less mRNAs produce that amount of proteins before degradation. Taking derivative of H with respect to d:

$$\frac{\partial H}{\partial d} = 8e^{-3dL} + 8de^{-3dL}(-3L) = 8e^{-3dL}(1-3dL).$$

Thus H is increasing with d if $d < \frac{1}{3L}$ otherwise is decreasing with d. Note that $\frac{1}{3L}$ is the reciprocal of t_c^{MC} and the average lifetime of an mRNA is $\frac{1}{d}$. Namely, if the average lifetime of an mRNA is shorter than t_c^{MC}, with the increase of d, more mRNAs produce proteins at the critical value and vice versa.

H tends to 0 as d, L goes to infinity:

$$\lim_{d,L\to\infty} H = 8d \lim_{d,L\to\infty} e^{-3dL} = 0.$$

Now we consider $f_N\left(C^+\right)$, which is given by:

$$f_N\left(C^+\right) = 4d \exp\left\{-d\left(4\cdot\frac{L}{4}+2L\right)\right\} = 4de^{-3dL} = \frac{H}{2}.$$

Hence, the probability density is halved as N increases through the critical point.

3.5 Discussion and Conclusion

In this chapter, we studied two typical mRNA degradation types and investigated how these degradation types control the amount of proteins produced from one mRNA during its lifetime. In our model, an mRNA in active translation is triggered to degrade by degradosome binding. For the tracking degradation, the degradosome tracks along and digests the mRNA. For the rapid degradation, the mRNA is degraded immediately after degradosome binding occurs.

We studied the probability density distribution for protein production of one mRNA. Table 3.1 lists the properties of the probability density distributions for the tracking degradation. One can use these properties to control the protein production. E represents the expected value of the protein production. To improve the average protein

		α	β	d	L
Low density phase	E	$\nearrow$	$-$	$\searrow_*$	$-$
Maximal current phase	E	$-$	$-$	$\searrow_*$	$-$
High density phase $\alpha < 1/2$	E	$\nearrow$	$\nearrow^*$	$\searrow_*$	$\nearrow^*$
	C	$\searrow_*$	$\nearrow$	$-$	$\nearrow$
	H	$\nearrow$	$\searrow_*$	$\curvearrowright_*$	$\searrow_*$
High density phase $\alpha \geqslant 1/2$	E	$-$	$\nearrow^*$	$\searrow_*$	$\nearrow^*$
	C	$-$	$\nearrow$	$-$	$\nearrow$
	H	$-$	$\searrow_*$	$\curvearrowright_*$	$\searrow_*$

Table 3.1: **Parameters determine the properties of probability density function for tracking degradation.** E is the expected value. C is the critical point at which the probability density function for high density phase is discontinuous. H is the probability density at the critical point. $-$: independent of the parameter. $\nearrow$: increasing with the parameter. $\searrow$: decreasing with the parameter. $\curvearrowright$: increasing first and then decreasing with the parameter. $*$: reaching a limiting level.

production for low density phase, for example, one can increase the ribosome binding rate, α, or decrease degradosome binding rate, d. Note that this is an intuitive result, in practice, it is not easy to change these parameters. C represents the critical point at which the probability density function is discontinuous and a peak occurs. H denotes the value of the probability density at C. The peak H indicates a higher likelihood of protein production around C. The discontinuity only occurs in the high density phase and is due to the change of the protein production rate during the translation process. H and C are negatively related. For example, for the high density phase when $\alpha < \frac{1}{2}$, increasing α gives a lower value of C and a higher likelihood H. In particular, C is independent of d, whilst H first increases and then decreases with d. Table 3.2 lists the properties of the probability density distributions for the rapid degradation. We introduced an extra quantity Z for this case. Z represents the probability of producing no protein, $P(0)$. Z is high compared to the rest of the probability densities, which means there is a high likelihood to produce no protein. There is a discontinuity in the low density phase and the maximal current phase. The value of the probability density, H, is halved as the probability density distribution passes through the critical point.

		α	β	d	L
Low density phase	E	$\curvearrowright$	$-$	$\searrow^*$	$\searrow^*$
	Z	$\searrow^*$	$-$	$\nearrow^*$	$\nearrow^*$
	C	$\nearrow$	$-$	$-$	$\nearrow$
	H	$\searrow^*$	$-$	$\curvearrowright^*$	$\searrow^*$
High density phase	E	$-$	$\nearrow$	$\searrow^*$	$\searrow^*$
	Z	$-$	$\searrow^*$	$\nearrow^*$	$\nearrow^*$
Maximal current phase	E	$-$	$-$	$\searrow^*$	$\searrow^*$
	Z	$-$	$-$	$\nearrow^*$	$\nearrow^*$
	C	$-$	$-$	$-$	$\nearrow$
	H	$-$	$-$	$\curvearrowright^*$	$\searrow^*$

Table 3.2: **Parameters determine the properties of probability density function for trapid degradation.** E is the expected value. Z is the probability of producing no protein. C is the critical point at which the probability density function for low density phase/maximal current phase is discontinuous. H is the probability density at the critical point. $-$: E, C, H is independent of the parameter. $\nearrow$: increasing with the parameter. $\searrow$: decreasing with the parameter. $\curvearrowright$: increasing first and then decreasing with the parameter. $*$: reaching a limiting level.

We studied the expected protein production and found that it is bounded by some limiting levels. We summarise the bounds as follows. For the tracking degradation:

$$\begin{aligned}
&\text{Low density phase:} \quad E[N] = \frac{\alpha(1-\alpha)}{d},\\
&\text{High density phase when } \alpha < \frac{1}{2}: \quad \frac{\beta(1-\beta)}{d} < E[N] < \frac{\alpha(1-\alpha)}{d},\\
&\text{High density phase when } \alpha \geqslant \frac{1}{2}: \quad \frac{\beta(1-\beta)}{d} < E[N] < \frac{1}{4d},\\
&\text{Maximal current phase:} \quad E[N] = \frac{1}{4}.
\end{aligned}$$

For the rapid degradation:

$$\begin{aligned}
&\text{Low density phase:} \qquad E[N] < \frac{\alpha(1-\alpha)}{d},\\
&\text{High density phase:} \qquad E[N] < \frac{\beta(1-\beta)}{d},\\
&\text{Maximal current phase:} \qquad E[N] < \frac{1}{4d}.
\end{aligned}$$

It is obvious that with the same parameter configuration, the average level of the protein production of the tracking degradation is higher than that of the rapid degradation for

all the phases. Comparing the average level of the protein production within the phases for the tracking degradation we see that:

①: given the same value of d and α, $\alpha < \frac{1}{2}$, $E_{HD} < E_{LD} < E_{MC}$,
②: given the same value of d, $\alpha \geqslant \frac{1}{2}$, $E_{LD} < E_{MC}$ and $E_{HD} < E_{MC}$,
③: for E_{LD} and E_{HD} when $\alpha \geqslant \frac{1}{2}$, values are dependent on β,

where the indices of E indicate the phases. Comparing the average level of the protein production within the phases for the rapid degradation we see that:

①: given the same value of d, $E_{LD} < E_{MC}$ and $E_{HD} < E_{MC}$,
②: for E_{LD} and E_{HD}, values are dependent on parameters.

Overall, our approximated probability density functions are in good agreement with the simulation results, and the probability density functions are able to reveal the basic properties of the mRNA degradation-translation model. However, the approximation for the rapid degradation in the maximal current phase need to be improved. Next, we will consider another mechanism to control the protein production: auto-negative feedback.

Chapter 4

The Role of Auto-Negative Feedback in Regulating Translation

4.1 Introduction

In some cases, gene expression is controlled by feedback at the translational level. For example, *E. coli* threonyl-tRNA synthetase represses its own mRNA translation by binding to the ribosome binding site to prevent binding of the ribosome [15, 20]. In this chapter, we consider a general case of protein feedback in translation. We study the situation where there is competition for the ribosome binding site on a given mRNA between ribosomes and the proteins produced from the mRNA itself. In some cases, proteins need to form into a complex before binding to the ribosome binding site. We assume that a protein complex binding/unbinding event is much faster and occurs more frequently than a ribosome binding/unbinding event. We construct a model based on the TASEP and use the Monte Carlo method to simulate this model. By applying the framework of the mean field solution of the TASEP and delay differential equation model, we gain fundamental insights about the biological process.

4.2 Model Construction and Simulation

4.2.1 Model Construction

Without loss of generality, we assume that a protein complex and a ribosome compete to bind to the vacant ribosome binding site. Due to the simple steric exclusion, they cannot bind simultaneously. Protein complex formation is required:

$$\underbrace{\mathrm{P}+\mathrm{P}+\ldots\mathrm{P}}_{n} \underset{r_2}{\overset{r_1}{\rightleftharpoons}} n\mathrm{P},$$

where P is protein, n proteins form a protein complex nP and r_1, r_2 are constant having dimension 1/time. In fact, protein multimerization evolves on a time scale that is much faster than that of translation initiation [23]. Therefore it is reasonable to apply a quasi-steady-state approximation on this reaction:

$$\frac{d[n\mathrm{P}]}{dt} = r_1\left([\mathrm{P}]\right)^n - r_2[n\mathrm{P}] \approx 0 \quad \Longrightarrow \quad [n\mathrm{P}] = \frac{r_1}{r_2}\left([\mathrm{P}]\right)^n = \frac{r_1}{r_2}\left(\frac{N}{V}\right)^n, \qquad (4.2.1)$$

$[\cdot]$ represents concentration and N denotes the number of molecules of P and V denotes the cell volume.

As shown in the schematic Figure 4.1, the model set up is as follows. A protein complex binds to the empty ribosome binding site at rate k_1. Law of mass action indicates:

$$k_1 = r_3[n\mathrm{P}] = \frac{r_1 r_3}{r_2}\left(\frac{N}{V}\right)^n, \qquad (4.2.2)$$

where r_3 is a constant with dimension 1/time. The protein complex is assumed to unbind from the ribosome binding site at rate k_2. A ribosome binds to the empty ribosome binding site at rate α. Ribosomes on site i, $i \in [1, L-1]$ jump to site $i+1$ at rate γ provided site $i+1$ is empty. Ribosome leaves site L at rate β, and a protein is

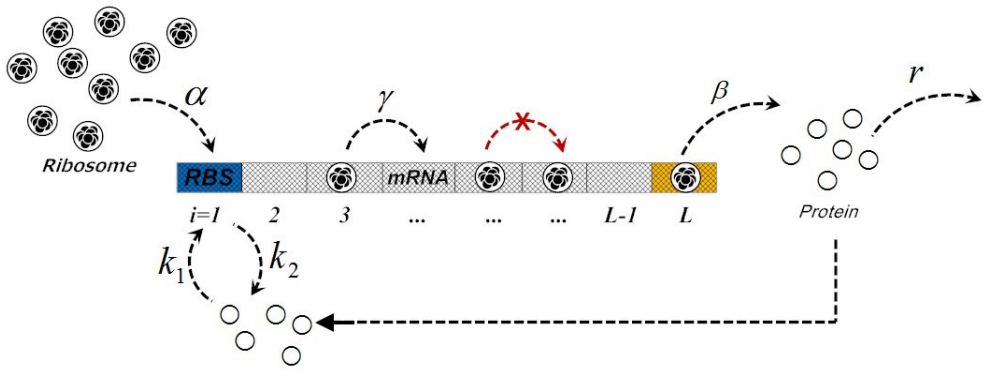


Figure 4.1: **Schematic of the model of translation with auto-negative feedback.** Protein is produced from the translation process and it negatively regulates its own translation. This model is based on the TASEP, please refer to the text for details. Unless otherwise noted, $\gamma = 1$.

produced. Proteins are removed at rate r to mimic all the possible losses.

4.2.2 Model Simulation

We use Monte Carlo method to simulate the model. The ribosome reservoir is indicated by site 0, the protein reservoir is indicated by site $L+1$. A Monte Carlo step is defined as $L+3$ 'attempts' (including protein complex binding/unbinding). At each attempt, a site i, $i \in [0, L+1]$ is randomly chosen to update. Particularly, $S_1 = 2$ indicates that site 1 is occupied by a protein complex. The algorithm is as follows.

1. Set $S_i = 0$, $i \in [1, L]$, protein particle number $N = 0$ and $t = 0$, set a MCS, τ,
2. Randomly choose a site:

   ①: If site 0 is chosen and $S_1 = 0$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \alpha\tau$, set $S_1 = 1$;

   ②: If site 1 is chosen, draw a uniform $(0, 1)$ random number ξ,

   ⅰ: If $S_1 = 1$ and $S_2 = 0$, set $S_1 = 0$, $S_2 = 1$;

   ⅱ: If $S_1 = 2$ and $\xi < k_2\tau$, set $S_1 = 0$;

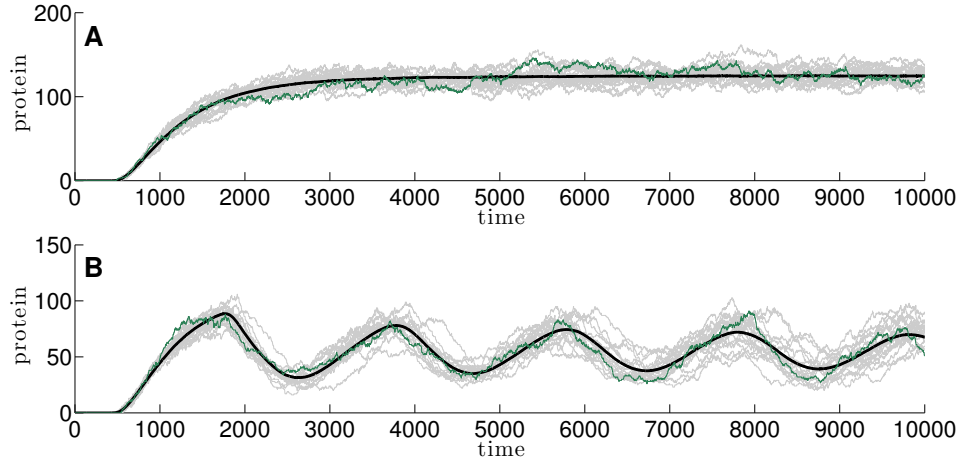


Figure 4.2: **Simulation of the model of translation with auto-negative feedback.** Protein number is plotted as a function of time. Black curves are the averages of 5000 stochastic simulations. 20 realizations of stochastic simulations are displayed for each case. Green curve is one of them. (A): Protein levels stabilise. $\alpha = 0.8$, $\beta = 0.8$, $V = 150$, $r_1/r_2 = 1$, $r_3 = 32$, $k_2 = 32$, $\tau = 0.025$. (B): Protein levels oscillate. $\alpha = 0.3$, $\beta = 0.7$, $V = 100$, $r_1/r_2 = 32$, $r_3 = 1$, $k_2 = 1$, $\tau = 0.025$. $\gamma = 1$, $L = 500$, $n = 5$, $r = 0.002$ for both cases. Initial protein levels are 0. Note that the trajectories fire at about 500. That is because it takes about L for the very first ribosome to produce a protein. Unless otherwise noted, this statement applies to all the remaining simulations.

③: If site i, $i \in [2, L-1]$ is chosen, and $S_i = 1$, $S_{i+1} = 0$, set $S_i = 0$, $S_{i+1} = 1$;

④: If site L is chosen and $S_L = 1$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \beta\tau$, set $S_L = 0$ and $N = N + 1$;

⑤: If site $L+1$ is chosen and $S_1 = 0$, draw a uniform $(0, 1)$ random number ξ: if $\xi < k_1\tau$, set $S_1 = 2$,

3. Repeat step 2 for $L+3$ times,

4. Set $c = 0$ (as a counter) and repeat the following procedure N times for protein removal: draw a uniform $(0, 1)$ random number ξ: if $\xi < r\tau$, set $c = c + 1$,

5. Set $N = N - c$ and $t = t + \tau$, repeat step $2-5$ or else stop the process.

Note that the protein multimerisation is not implemented explicitly in the simulation, since the quasi-steady-state approximation for that has been made to simplify the

whole process. Using the Monte Carlo method we introduced above, we obtain typical simulations as shown in Figure 4.2. Note that in the simulation, time is rescaled, 10000 corresponds to about 5.5 minutes in real time (detail is shown in Appendix B). After extensive parameter testing, two typical cases are found: (i) protein levels stabilise (see Figure 4.2A), (ii) protein levels oscillate (see Figure4.2B). In general, negative feedback and time delay can create an oscillatory response [69, 134]. In our system, protein negatively regulates translation initiation, and the ribosome transportation in the translation process naturally provides a time delay. Thus oscillations are not unexpected. However, the competition between protein and ribosome on the ribosome binding site makes a detailed mathematical analysis difficult. Later on, we will make some assumptions to make a simple analysis for a basic understanding of the model dynamics.

4.3 Simplified Model: Equilibrium Approximation

4.3.1 Model Formulation

We now consider the competition on the ribosome binding site which is expressed by the following biochemical reactions:

$$n\mathrm{P} + \mathrm{RBS} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} n\mathrm{P}\cdot\mathrm{RBS}, \tag{4.3.1}$$

$$\mathrm{R} + \mathrm{RBS} \xrightarrow{\alpha} \mathrm{R}\cdot\mathrm{RBS} \xrightarrow{1-S_2} \mathrm{RBS}, \tag{4.3.2}$$

where $n\mathrm{P}\cdot\mathrm{RBS}$ is the complex of the ribosome binding site and the protein complex, RBS is the vacant ribosome binding site and $\mathrm{R}\cdot\mathrm{RBS}$ is the complex of ribosome and the ribosome binding site. Assume the biochemical reaction (4.3.1) occurs rapidly and

is dominant, $n\text{P}\cdot\text{RBS}$ is in instantaneous equilibrium, i.e., k_1, k_2 are large compared to α, $1-S_2$, then we can derive the following equilibrium approximations:

$$\frac{d[n\text{P}\cdot\text{RBS}]}{dt} = k_1 \times [n\text{P}] \times [\text{RBS}] - k_2 \times [n\text{P}\cdot\text{RBS}] \approx 0. \tag{4.3.3}$$

Since the reaction (4.3.1) occurs much more frequently and faster than the reaction (4.3.2), we assume that (4.3.1) and (4.3.2) can be regarded as two separate independent processes: (i) nP binds/unbinds to the RBS frequently, leading to a probability of the RBS being vacant (not occupied by nP), and (ii) ribosome binds to the vacant RBS. For process (i), the ribosome binding site is either occupied by a protein complex or is vacant for ribosome to bind. Let us consider a single mRNA. The total concentration of ribosome binding site is denoted by $1/V$, where V is the cell volume. Therefore, $[\text{RBS}] + [n\text{P}\cdot\text{RBS}] = 1/V$. From (4.3.3) and using the fact (4.2.1):

$$\begin{aligned} k_1 \times [n\text{P}] \times [\text{RBS}] \approx k_2 \times [n\text{P}\cdot\text{RBS}] &= k_2 \left(\frac{1}{V} - [\text{RBS}] \right) \\ \Longrightarrow [\text{RBS}] \approx \frac{V^{-1}}{\left([\text{P}] \Big/ \sqrt[n]{r_2 k_2 / r_1 k_1} \right)^n + 1} &= \frac{V^{-1}}{1 + \left(N/\theta \right)^n}, \end{aligned}$$

by setting $\theta = V \sqrt[n]{r_2 k_2 / r_1 k_1}$ and the number of protein N to be $[\text{P}] \times V$. Let $f(N)$ denote the probability that the ribosome binding site is free:

$$p(N) = \frac{[\text{RBS}]}{V^{-1}} = \frac{V^{-1} / \left(1 + \left(N/\theta \right)^n \right)}{V^{-1}} = \frac{1}{1 + \left(N/\theta \right)^n}.$$

Therefore, considering the process (4.3.2), we define the effective ribosome binding rate as the following Hill function:

$$\alpha_{eff} = \alpha \times p(N) = \frac{\alpha}{1 + \left(N/\theta \right)^n}. \tag{4.3.4}$$

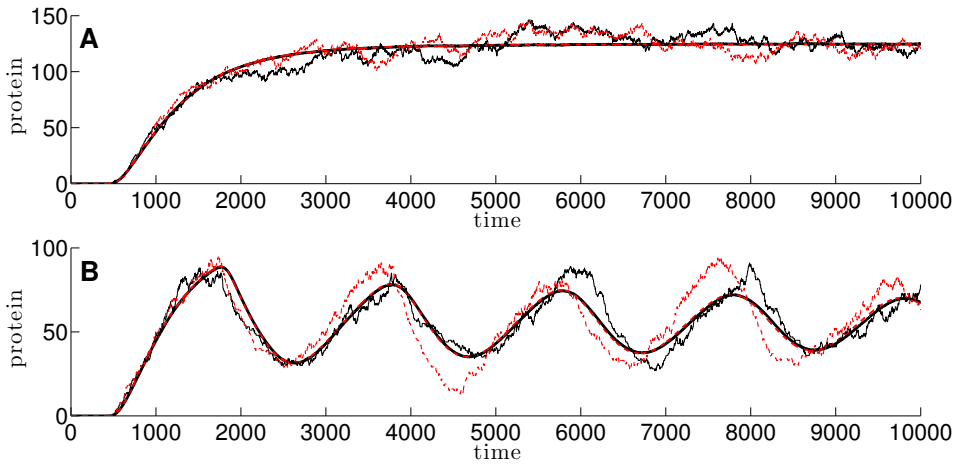


Figure 4.3: **Simulation test for the simplifed model.** Protein number is plotted as a function of time. Black curves are copies of what we showed in Figure 4.2. Red dashed curves are the average of 5000 stochastic simulations of the simplifed model. One realisation of simulation is displayed for each case. We use the same parameter configuration as what we used for Figure 4.2. (A): Protein levels stabilise. $\alpha = 0.8$, $\beta = 0.8$, $\theta = 150$. (B): Protein levels oscillate. $\alpha = 0.3$, $\beta = 0.7$, $\theta = 50$. $L = 500$, $n = 5$, $r = 0.002$ for both cases. Intial protein levels are 0.

Now we are able to simplify the model by considering such a TASEP: ribosomes bind to the empty ribosome binding site at rate α_{eff}, ribosomes on the lattice jump to the next site provided the next site is empty, ribosomes finally leave site L and a protein is produced, proteins are removed at rate r. In essence, in the simplified model, the protein complex formation and the binding/unbinding of the protein complex to the ribosome binding site are incorporated into α_{eff}. Before we perform the mean field analysis, we do a simulation test to check the reasonability of the assumptions. The simulation test is shown in Figure 4.3. As we can see, the simulation of the simplified model and that of the original auto-negative feedback model match well and no significant differences of the fluctuations are found. Thus, we are confident that the assumptions in the simplified model are reasonable. The simplified model is well developed to reveal the properties of the original model. Next we perform the mean field analysis by using α_{eff} and β. The mean field approach is mainly modified in the low density phase since the properties of the low density phase is determined by α_{eff}.

Generalised phase boundaries will be obtained in Section 4.3.2-4.3.4. Further discussion on these results will be made later in Section 4.3.5. Note that in the steady state, the mean field protein level is given by J/r, which can be obtained by considering the steady state solution of the following ODE:

$$\frac{dN(t)}{dt} = J - rN(t).$$

4.3.2 High Density Phase

Knowing that $J_{HD} = \beta(1-\beta)$ and $N = \frac{J_{HD}}{r}$, together with (4.3.4), we have:

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n} = \frac{\alpha}{1+\left(\frac{\beta(1-\beta)}{\theta r}\right)^n}. \tag{4.3.5}$$

The simplified model is in the high density phase if $\alpha_{eff} > \beta, \beta < \frac{1}{2}$. Hence, substituting (4.3.5) into these inequalities yields the boundary for the high density phase:

$$\alpha > \left(1+\left(\frac{\beta(1-\beta)}{\theta r}\right)^n\right)\beta, \qquad \beta < \frac{1}{2}. \tag{4.3.6}$$

In the high density phase, $J_{HD} = \beta(1-\beta), \rho_{HD} = 1-\beta$, the auto-negative feedback does not affect the properties at all. The protein levels stabilise.

4.3.3 Maximal Current Phase

Knowing that $J_{MC} = \frac{1}{4}$ and $N = \frac{1}{4r}$, together with (4.3.4), we have:

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n} = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}. \tag{4.3.7}$$

The simplified model is in the maximal current phase if $\alpha_{eff},\ \beta \geqslant \frac{1}{2}$. Hence, substituting (4.3.7) into these inequalities yields the maximal current phase boundary:

$$\alpha \geqslant \frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^n\right), \qquad \beta \geqslant \frac{1}{2}. \tag{4.3.8}$$

In the maximal current phase, $J_{MC}=\frac{1}{4}$, $\rho_{MC}=\frac{1}{2}$, the auto-negative feedback does not affect the properties at all. The protein levels stabilise.

4.3.4 Low Density Phase

The simplified model is in the low density phase if:

$$\alpha_{eff} < \frac{1}{2}, \qquad \beta > \alpha_{eff}. \tag{4.3.9}$$

Knowing that $J_{LD}=\alpha_{eff}\left(1-\alpha_{eff}\right)$ and $N=\frac{J_{LD}}{r}<\frac{1}{4r}$, we have:

$$\alpha_{eff}=\frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n}=\frac{\alpha}{1+\left(\frac{\alpha_{eff}\left(1-\alpha_{eff}\right)}{\theta r}\right)^n}. \tag{4.3.10}$$

In the low density phase, α_{eff} is the solution of (4.3.10), $J_{LD}=\alpha_{eff}\left(1-\alpha_{eff}\right)$, $\rho_{LD}=\alpha_{eff}$. However, the solutions of (4.3.10) is not obvious, we have to ensure the solutions exist. Multiplying $(\theta r)^n+\left(\alpha_{eff}(1-\alpha_{eff})\right)^n$ at both sides yields:

$$f=\alpha_{eff}^{n+1}\left(1-\alpha_{eff}\right)^n+\alpha_{eff}(\theta r)^n-\alpha(\theta r)^n=0. \tag{4.3.11}$$

The derivative of f with respect to α_{eff} is:

$$f'=(n+1)\alpha_{eff}^n\left(1-\alpha_{eff}\right)^n-n\alpha_{eff}^{n+1}\left(1-\alpha_{eff}\right)^{n-1}+(\theta r)^n$$

$$= \alpha_{eff}^{n}\left(1-\alpha_{eff}\right)^{n-1}\left(1-\alpha_{eff}+n(1-2\alpha_{eff})\right)+(\theta r)^{n},$$

since both $1-\alpha_{eff}$ and $(1-2\alpha_{eff})$ are positive in $\left(0,\frac{1}{2}\right)$, so $f'>0$ in $\left(0,\frac{1}{2}\right)$, namely, f is increasing. Moreover, $f(0)=-\alpha(\theta r)^{n}<0$. From (4.3.9), $\alpha_{eff}<\min\left(\frac{1}{2},\beta\right)$. Therefore, the solution of (4.3.10) under condition (4.3.9) exists if and only if:

$$f\left(\min\left(\frac{1}{2},\beta\right)\right)>0, \tag{4.3.12}$$

and the solution is unique.

For $\beta<\frac{1}{2}$, $f(\beta)>0$ yields:

$$f(\beta)=\beta^{n+1}(1-\beta)^{n}+\beta(\theta r)^{n}-\alpha(\theta r)^{n}>0,$$

which gives:

$$\alpha<\beta\left(1+\left(\frac{\beta(1-\beta)}{\theta r}\right)^{n}\right). \tag{4.3.13}$$

For $\beta\geqslant\frac{1}{2}$, $f\left(\frac{1}{2}\right)>0$ yields:

$$f\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{n+1}\left(1-\frac{1}{2}\right)^{n}+\frac{1}{2}(\theta r)^{n}-\alpha(\theta r)^{n}>0,$$

which gives:

$$\alpha<\frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right). \tag{4.3.14}$$

In summary, the boundary for the low density phase is:

$$\begin{cases}\alpha<\dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^{n}\right) & \beta\geqslant\dfrac{1}{2},\\[2ex] \alpha<\beta\left(1+\left(\dfrac{\beta(1-\beta)}{\theta r}\right)^{n}\right) & \beta<\dfrac{1}{2}.\end{cases} \tag{4.3.15}$$

We have known that due to the time delay and the negative feedback, oscillatory behaviour can arise (see Figure 4.3B). The above mean field analysis is incapable of clarifying the oscillation (mean field only gives the steady state solution and ignores the correlation between sites). We now develop a delay differential equation (DDE) model to study the oscillation in the low density phase. Assume that ribosomes require a period of time T to travel from the 5' end to the 3' end of an mRNA. Thus, the protein production rate at time t is determined by the rate of ribosomes entering the mRNA at time $t-T$, i.e., $J(t) = \alpha_{eff}\left(N(t-T)\right)\left(1-\alpha_{eff}\left(N(t-T)\right)\right)$. Thus, the DDE model is described by:

$$\frac{dN(t)}{dt} = \frac{\alpha}{1+\left(\frac{N(t-T)}{\theta}\right)^n}\left(1-\frac{\alpha}{1+\left(\frac{N(t-T)}{\theta}\right)^n}\right) - rN(t). \tag{4.3.16}$$

The steady state of (4.3.16), N^*, is the solution of the following equation:

$$R(N) = \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n}\left(1-\frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n}\right) - rN = 0. \tag{4.3.17}$$

We now check if N^* exists within $\left(0, \frac{1}{4r}\right)$. Taking derivative of $R(N)$ with respect to N:

$$\begin{aligned}\frac{dR(N)}{dN} &= \left(1-\frac{2\alpha}{1+\left(\frac{N}{\theta}\right)^n}\right)\frac{d}{dN}\left(\frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n}\right) - r \\ &= \left(1-\frac{2\alpha}{1+\left(\frac{N}{\theta}\right)^n}\right)\frac{-\alpha n\left(\frac{N}{\theta}\right)^n}{N\left(1+\left(\frac{N}{\theta}\right)^n\right)^2} - r.\end{aligned}$$

In the low density phase, $\frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n} < \frac{1}{2} \Longrightarrow 1 - \frac{2\alpha}{1+\left(\frac{N}{\theta}\right)^n} > 0$. Letting

$$B = \left(1 - \frac{2\alpha}{1+\left(\frac{N}{\theta}\right)^n}\right) \frac{\alpha n \left(\frac{N}{\theta}\right)^n}{N\left(1+\left(\frac{N}{\theta}\right)^n\right)^2} > 0, \tag{4.3.18}$$

then $\frac{dR(N)}{dt} = -B - r < 0$, $R(N)$ is a monotone decreasing function of N. Moreover, $R(0) = \alpha(1-\alpha) > 0$ and

$$R\left(\frac{1}{4r}\right) = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\left(1 - \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right) - r\frac{1}{4r} < \frac{1}{4} - \frac{1}{4} < 0.$$

Therefore, unique N^* exists. Investigating the oscillatory behaviour requires stability analysis.

To check the stability of N^*, we consider a small perturbation $|\delta(t)| \ll 1$. Substituting $N(t) = N^* + \delta(t)$ into (4.3.16) and using Taylor expansion (the high order term is dropped) yields:

$$\frac{d(N^*+\delta(t))}{dt} = \frac{\alpha}{1+\left(\frac{N^*+\delta(t-T)}{\theta}\right)^n}\left(1 - \frac{\alpha}{1+\left(\frac{N^*+\delta(t-T)}{\theta}\right)^n}\right) - r\left(N^*+\delta(t)\right),$$

$$\frac{d\delta(t)}{dt} = \frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n}\left(1 - \frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n}\right) - rN^* - B\delta(t-T) - r\delta(t) + O\left(\delta^2\right),$$

$$\frac{d\delta(t)}{dt} = -B\delta(t-T) - r\delta(t), \tag{4.3.19}$$

where $B > 0$ and B is given by (4.3.18). Suppose that $\delta(t)$ has a solution $\delta(t) = Ae^{\lambda t}$

and substitute this solution into (4.3.19):

$$\lambda + r = -Be^{-\lambda T}. \tag{4.3.20}$$

Any real solution λ of (4.3.20) is negative, since $\lambda = -(r + Be^{-\lambda T}) < 0$. Suppose that when $T = T_c$, the purely imaginary eigenvalues arise, i.e., $\lambda_c = \pm i\omega$, a Hopf bifurcation could occur, note that this is a necessary but not sufficient condition [73]. We suppose that ω is real and positive. Given that $e^{ix} = \cos x + i \sin x$, substituting $\lambda = \pm i\omega$ into (4.3.20) yields:

$$\pm i\omega + r = -Be^{\mp i\omega T_c} \quad \Longrightarrow \quad \begin{cases} r = -B\cos\omega T_c \\ \omega = B\sin\omega T_c \end{cases}. \tag{4.3.21}$$

From (4.3.21), it follows that $\omega^2 = B^2(\sin\omega T_c)^2 = B^2 - B^2(\cos\omega T_c)^2 = B^2 - r^2$. So $\omega = \sqrt{B^2 - r^2}$ provided $B^2 - r^2 > 0$. Since $B^2 - r^2 = (B + r)(B - r)$, and $B,\ r > 0$, the sign of $B^2 - r^2$ is determined by $B - r$. Using (4.3.17), (4.3.18), we are able to ascertain that:

$$\begin{aligned} &B - r \\ &= \left(1 - \frac{2\alpha}{1 + \left(\frac{N^*}{\theta}\right)^n}\right) \frac{\alpha n \left(\frac{N^*}{\theta}\right)^n}{N\left(1 + \left(\frac{N^*}{\theta}\right)^n\right)^2} - \frac{\alpha}{N^*\left(1 + \left(\frac{N^*}{\theta}\right)^n\right)} \left(1 - \frac{\alpha}{1 + \left(\frac{N^*}{\theta}\right)^n}\right) \\ &= \frac{\alpha\left((n-1)\left(\frac{N^*}{\theta}\right)^{2n} + (\alpha + n - 2\alpha n - 2)\left(\frac{N^*}{\theta}\right)^n + \alpha - 1\right)}{N^*\left(1 + \left(\frac{N^*}{\theta}\right)^n\right)^3}, \end{aligned}$$

We define:

$$g\left(\left(\frac{N^*}{\theta}\right)^n\right) = (n-1)\left(\frac{N^*}{\theta}\right)^{2n} + (\alpha + n - 2\alpha n - 2)\left(\frac{N^*}{\theta}\right)^n + \alpha - 1. \tag{4.3.22}$$

Hence, the sign of $B-r$ is determined by $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$, which is a quadratic function.

First, we notice that if $n=1$, then $g\left(\left(\frac{N^*}{\theta}\right)^n\right) = -(\alpha+1)\frac{N^*}{\theta} - (1-\alpha) < 0$. Hence a Hopf bifurcation cannot occur since we require $B-r>0$. Therefore, one requirement for a Hopf bifurcation is that $n>1$.

For $n>1$, the discriminant of $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$ is given by:

$$\Delta = (\alpha+n-2\alpha n-2)^2 + 4(n-1)(1-\alpha) > 0, \tag{4.3.23}$$

since $\alpha<1$. $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$ has the following two distinct real roots:

$$\left(\frac{N^*}{\theta}\right)^n_{\pm} = \frac{-(\alpha+n-2\alpha n-2) \pm \sqrt{\Delta}}{2(n-1)},$$

where Δ is given in (4.3.23). However, $\left(\frac{N^*}{\theta}\right)^n_-$ is negative since:

$$\left(-(\alpha+n-2\alpha n-2)\right)^2 - \Delta = -4(n-1)(1-\alpha) < 0.$$

Therefore, it is easy to conclude that $g\left(\left(\frac{N^*}{\theta}\right)^n\right) > 0$ requires:

$$\left(\frac{N^*}{\theta}\right)^n > \left(\frac{N^*}{\theta}\right)^n_+ = \frac{-(\alpha+n-2\alpha n-2)+\sqrt{\Delta}}{2(n-1)}. \tag{4.3.24}$$

From (4.3.17) and (4.3.24), it follows that:

$$\frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n_+}\left(1-\frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n_+}\right) > \frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n}\left(1-\frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)^n}\right) > r\theta\sqrt[n]{\left(\frac{N^*}{\theta}\right)^n_+}.$$

Hence, we get the following condition:

$$\theta r < \frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)_+^n}\left(1-\frac{\alpha}{1+\left(\frac{N^*}{\theta}\right)_+^n}\right)\frac{1}{\sqrt[n]{\left(\frac{N^*}{\theta}\right)_+^n}} \tag{4.3.25}$$

The condition (4.3.25) divides the low density phase into two regions. In one region, the condition (4.3.25) does not hold, there is no Hopf bifurcation for the DDE (4.3.16), N^* is stable, resulting in a stabilised protein level. In the other region, the condition (4.3.25) is satisfied and a Hopf bifurcation for the DDE (4.3.16) could arise, resulting in a oscillatory protein level. From (4.3.21) and given $\omega = \sqrt{B^2 - r^2}$, we are able to obtain T_c:

$$T_c = \frac{1}{\sqrt{B^2 - r^2}} \arcsin \frac{\sqrt{B^2 - r^2}}{B}. \tag{4.3.26}$$

Therefore, for $T \geqslant T_c$, oscillation could occur. T is used to approximate the ribosome travel time from one end of a mRNA to the other end, which is given by $T \approx \frac{L}{J}$, where J is the current. Thus, to obtain oscillatory behaviours, the system requires $\frac{L}{J} \geqslant T_c$, i.e.,

$$L \geqslant JT_c. \tag{4.3.27}$$

Otherwise noted, all the simulations and analysis in this thesis are satisfied the condition that $L \geqslant JT_c$.

4.3.5 Simulation and Discussion

We will show the simulations in this section. Recall that the good agreement of the simulations for the simplified model and the simulations for the original model (see Figure 4.3), we only display the simulations for the simplified model next. From the previous section, we summarise the mean field phase boundaries for the simplified

model as follows:

$$\begin{cases} \text{LD - HD:} \quad \alpha = \left(1+\left(\dfrac{\beta(1-\beta)}{\theta r}\right)^n\right)\beta, \quad \beta < \dfrac{1}{2}, \\ \text{LD - MC:} \quad \alpha = \dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right), \quad \beta \geqslant \dfrac{1}{2}, \\ \text{HD - MC:} \quad \beta = \dfrac{1}{2}, \quad \alpha \geqslant \dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right). \end{cases} \tag{4.3.28}$$

The boundaries (4.3.28) intersect at the point $(\alpha,\beta) = \left(\dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right), \dfrac{1}{2}\right)$. The maximal current phase vanishes for $\theta r < 1/4$. The boundary $\alpha = \dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right)$ is a straight horizontal line since n, θ, r are constants, and $\alpha = \dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right) > \dfrac{1}{2}$.

$$\begin{aligned} \because \quad \frac{d}{d\beta}\left(\beta\left(1+\left(\frac{\beta(1-\beta)}{r\theta}\right)^n\right)\right) &= 1+\frac{(n+1)\beta^n(1-\beta)^n - n\beta^{n+1}(1-\beta)^{n-1}}{(r\theta)^n} \\ &= 1+\frac{\beta^n(1-\beta)^{n-1}\overbrace{\left(n(1-2\beta)+(1-\beta)\right)}^{>0,\quad \beta\in[0,0.5]}}{(r\theta)^n} > 1 > 0, \end{aligned}$$

therefore the boundary $\alpha = \left(1+\left(\dfrac{\beta(1-\beta)}{\theta r}\right)^n\right)\beta, \beta < \dfrac{1}{2}$ is monotonically increasing, and $\alpha = \left(1+\left(\dfrac{\beta(1-\beta)}{\theta r}\right)^n\right)\beta > \beta$.

Recall that the boundaries for the standard TASEP is given as follows:

$$\begin{cases} \alpha = \beta, & \beta \leqslant \dfrac{1}{2}, \\ \alpha = \dfrac{1}{2}, & \beta > \dfrac{1}{2}, \\ \beta = \dfrac{1}{2}, & \alpha \geqslant \dfrac{1}{2}. \end{cases} \tag{4.3.29}$$

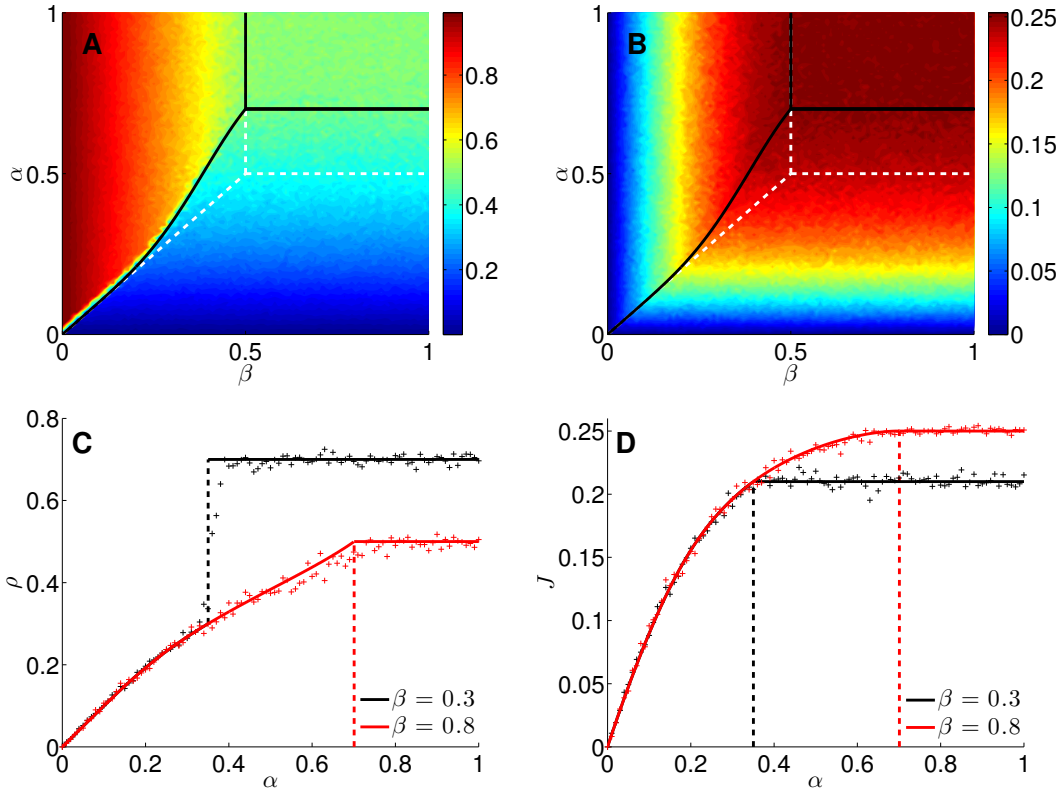


Figure 4.4: **Phase diagram and phase transition of the simplified model with no oscillation.** (A): Average density ρ of the simplified model. (B) Current J of the simplified model. Blacked lines are the mean field boundaries (4.3.28). White dashed lines are the boundaries of the TASEP (4.3.29). Phase transitions for the (C) average density and (D) current by fixing $\beta = 0.3$ or $\beta = 0.8$ and varying α systematically (dashed lines indicate the boundaries). Lines are mean field solutions, '+' labels the data from simulations. Parameters: $n = 5$, $r = 0.002$, $L = 500$, $\theta = 150$.

Thus compare (4.3.28) and (4.3.29), we can see that the boundaries for the simplified model are shifted and the low density phase is enlarged. Moreover, θr has a significant impact on the shape of (4.3.28). It is not difficult to see that as θr increases, the boundaries (4.3.28) degenerates to the boundaries (4.3.29) eventually as $\big(1/(\theta r)\big)^n$ goes to 0. In this sense, we name $\dfrac{1}{\theta r}$ *'Impact Factor'*. The impact factor measures the significance of the feedback. The larger the impact factor, the greater the effect the feedback has on the system. Interestingly, from (4.3.25), the impact factor also determines the boundary for the stable and oscillatory cases.

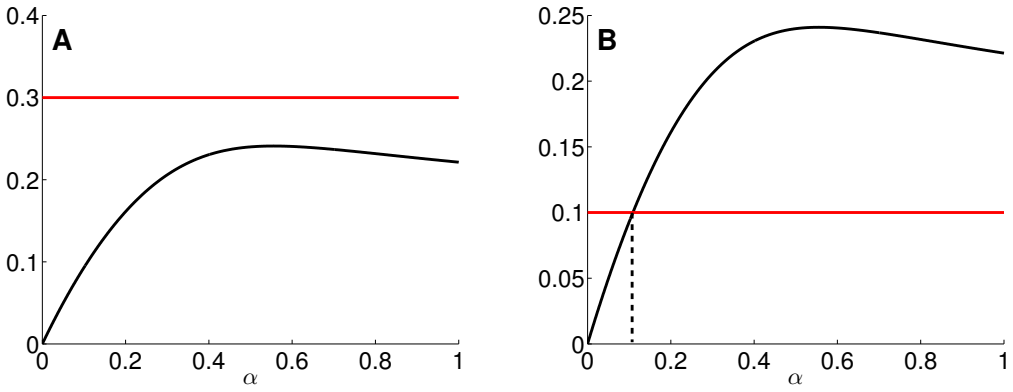


Figure 4.5: **Plotting condition (4.3.25) as a function of α.** Red line is θr. Black curve is the right hand side of (4.3.25). (A) Parameter configuration that cannot lead to oscillation, $\theta = 150,\ r = 0.002,\ n = 5$. (B) Parameter configuration that can lead to oscillation, $\theta = 50,\ r = 0.002,\ n = 5$

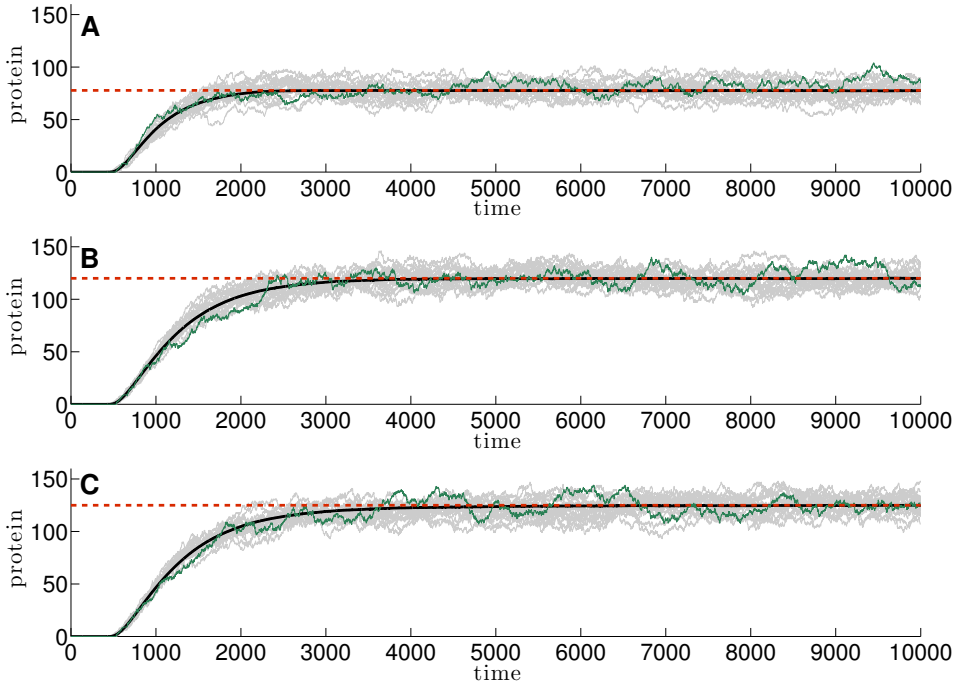


Figure 4.6: **Stochastic simulations of the simplified model: protein levels stabilise.** Protein number is plotted as a function of time. Black lines are the averages of 5000 realizations of stochastic simulations. Grey lines are 20 realizations of stochastic simulations, green line is one of them. Red dashed lines are mean field N^*. (A): LD. N^* is calculated from $\alpha_{eff}\left(1-\alpha_{eff}\right)/r$. $\alpha = 0.2, \beta = 0.6$. (B): HD. $N^* = \beta(1-\beta)/r$. $\alpha = 0.8, \beta = 0.4$. (C): MC. $N^* = 1/4r$. $\alpha = 0.8, \beta = 0.8$. All other parameters: $\theta = 150, n = 5, r = 0.002, L = 500$. The simulations are generated by the Monte Carlo method.

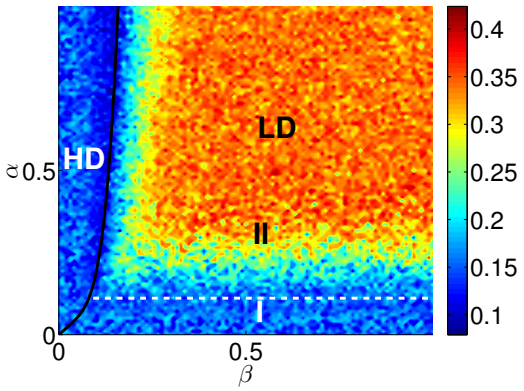


Figure 4.7: **Phase diagram for the simplified model with oscillation in the low density phase.** We plot C_v (4.3.30) in the plane. Under the chosen parameter configuration, only HD and LD phases exist in the plane $\alpha, \beta \in [0, 1]$. Black curve is the boundary between the low density phase and high density phase. White dashed line divides the low density phase into two parts ($\alpha = 0.1089$, obtained from Figure 4.5B). I: protein level stabilises. II: protein level oscillates. $L = 500$, $n = 5$, $\theta = 50$, $r = 0.002$.

In Figure 4.4A/B, we show a simulation of the phase diagram for the simplified model. We can see that the boundaries (black lines) are shifted compared to the boundaries for the standard TASEP (white lines). The low density phase is extended at the expense of the high density phase and the maximal current phase. The simulation results do show different behaviours in different phases. By fixing β and varying α systematically, the system undergoes LD-HD and LD-MC phase transitions. The phase transitions are shown in Figure 4.4C/D. The simulation agrees with the mean field solution well. In this example, the oscillation cannot occur. From Figure 4.5A, we can see that given the specific parameter configuration, the condition (4.3.25) can never be satisfied. We picked a set of parameter configurations from each phase shown in Figure 4.4, and obtained typical simulations for the simplified model. The simulation results are shown in Figure 4.6. The average of stochastic simulations reaches the mean field approximation eventually. The protein levels stabilise.

However, we have known that in the low density phase, oscillation can occur. For example, given the parameter configuration $\theta = 50$, $r = 0.002$, $n = 5, L = 500$, for

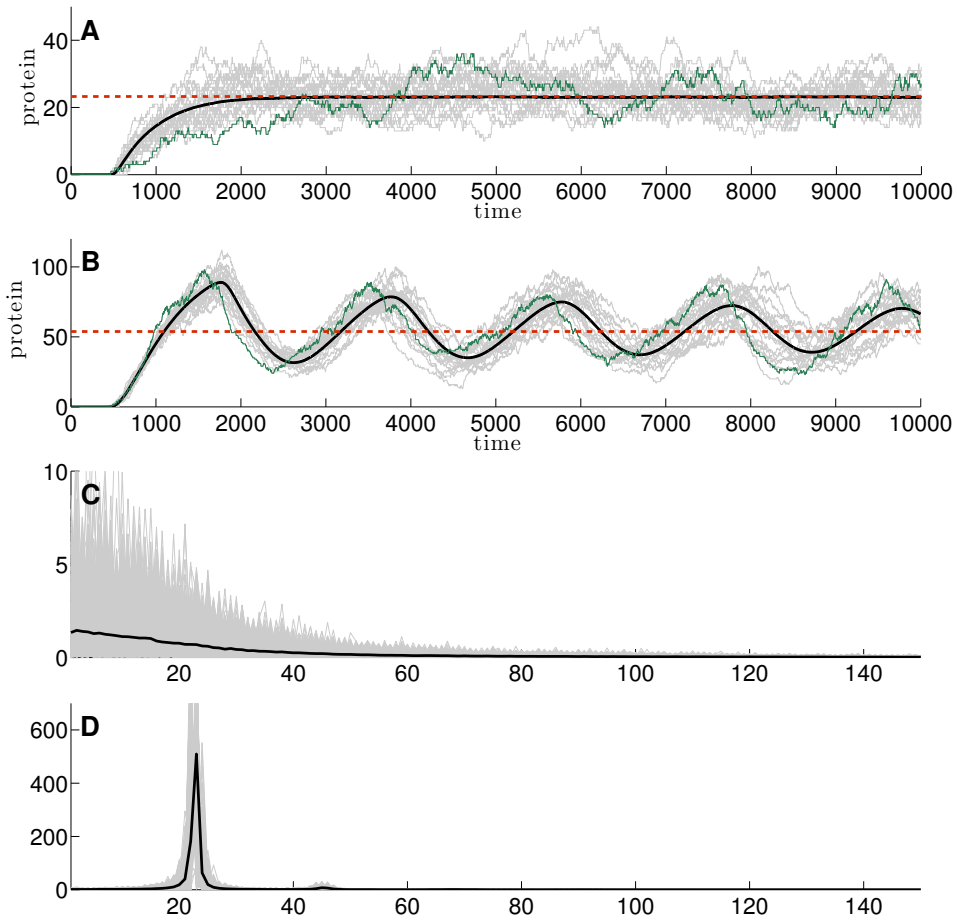


Figure 4.8: **Simulation and power spectrum for the simplified model in the low density phase.** Protein number is plotted as a function of time. Black lines are the averages of the 5000 realizations of stochastic simulations. Grey lines are 20 realizations of stochastic simulations, green line is one of them. Red dashed lines are mean field N^*. Typical simulation (A) and the corresponding power spectrum (C) from the region I shown in Figure 4.7. $\alpha = 0.05$, $\beta = 0.7$, $n = 5$, $\theta = 50$, $r = 0.002$, $L = 500$. Typical simulation (B) and the corresponding power spectrum (D) from the region II shown in Figure 4.7. $\alpha = 0.5$, $\beta = 0.7$, $n = 5$, $\theta = 50$, $r = 0.002$, $L = 500$.

$\alpha > 0.1089$, the condition for Hopf bifurcation (4.3.25) is satisfied (see Figure 4.5B). Thus, oscillation occurs when $\alpha > 0.1089$. We show the phase diagram under this parameter configuration in Figure 4.7. We used a standardised quantity, *coefficient of variation*, to measure the dispersion of the simulation. The coefficient of variation is defined as:

$$C_v = \frac{\sigma}{\mu}, \tag{4.3.30}$$

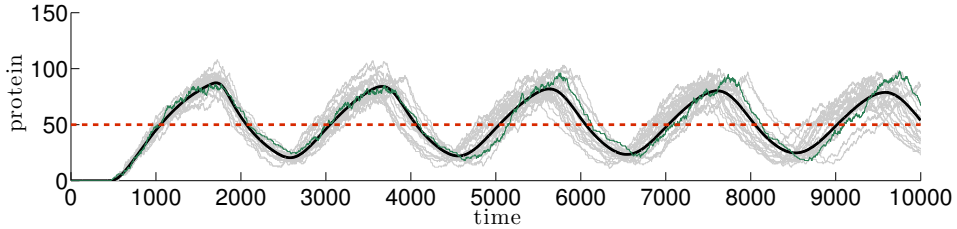


Figure 4.9: **Stochastic simulations for the limiting case.** Grey lines are 20 realizations of the stochastic simulations and green line is one of them. Black curve is the average of the 5000 realizations of stochastic simulations. Red dashed line is $y = \theta$. $\alpha = 0.3$, $\beta = 0.7$, $\theta = 50$, $L = 500$, $r = 0.002$.

where σ is the standard deviation, and μ is the mean of the protein level. Intuitively, C_v for oscillatory simulation is higher than C_v for stable simulation. From Figure 4.7, we do observe that $\alpha = 0.1089$ divides the low density phase into two regions. In region I, C_v is low. In region II, C_v is high. In Figure 4.8, we show typical simulation and corresponding power spectrum from the region I and II. From Figure 4.8A/C, we observe that in region I, the average of simulations tends to a stable level which given by the mean field solution and the corresponding power spectrum shows a broad band. Thus, in region I, protein levels stabilise. From 4.8B/D, we observe that in region II, the average of simulations displays a oscillatory behaviour and the corresponding power spectrum gives peaks. Thus, in region II, protein levels oscillate.

4.4 A Limiting Case

In the previous section, we affirmed the existence of the oscillation in the low density phase. However, there is very little we can do to determine the period and amplitude of the oscillation. In this section, we consider a limiting case of $n \to \infty$. We are able to write down exact expressions of the period and amplitude for the limiting model, then to provide a qualitative understanding for the general model.

As $n \to \infty$, the effective ribosome binding rate (4.3.4) can be approximated by:

$$\alpha_{eff} = \alpha H(\theta - N), \tag{4.4.1}$$

where $H(x)$ is the Heaviside function. A typical oscillation example is shown in Figure 4.9. In this limiting case, after the protein level passes through θ, the translation initiation is switched off/on. However, it takes some time for the ribosome to undergo the elongation process and the termination. Thus, a time delay T is introduced for the system to react to the switching off/on of the translation initiation. The delay $T \propto L$, where L is the length of the lattice. Particularly, $S \propto \alpha(1-\alpha)$ since the oscillation only occurs in the low density phase. In order to perform mathematical analysis, we develop the following simple deterministic DDE model:

$$\frac{dN(t)}{dt} = SH\left(\theta - N(t-T)\right) - rN(t). \tag{4.4.2}$$

Next we attempt to obtain the period and the amplitude of the oscillation in (4.4.2). Starting with $N(0) = 0$, $N(t)$ increases with time. Suppose that at time t_{c1}, $N(t_{c1}) = \theta$. Then $H\left(\theta - N(t-T)\right) = 1$ during $t \in [0, t_{c1} + T]$. The following ODE:

$$\frac{dN(t)}{dt} = S - rN(t), \qquad N(0) = 0, \tag{4.4.3}$$

has the solution: $N(t) = S\left(1 - e^{-rt}\right)/r$. Thus, $N(t_{c1}) = S\left(1 - e^{-rt_{c1}}\right)/r = \theta$, and we have $t_{c1} = -\ln\left(1 - \theta r/S\right)/r$.

Let $T_0 = t_{c1} + T$, then the protein number at $t = T_0$ is $N(T_0) = \frac{S}{r}\left(1 - e^{-rT_0}\right)$. Since $N(T_0) > \theta$, so after that, $H\left(\theta - N(t-T)\right) = 0$ and the protein number starts decreasing. Suppose that at time t_{c2}, $N(t_{c2}) = \theta$ and:

$$\frac{dN(t)}{dt} = -rN(t), \quad t \in (T_0, t_{c2} + T]. \tag{4.4.4}$$

The solution of (4.4.4) is $N(t) = N(T_0)e^{-r(t-T_0)}$. We have:

$$N(t_{c2}) = N(T_0)e^{-r(t_{c2}-T_0)} = \theta, \quad \Longrightarrow \quad t_{c2} = \frac{1}{r}\ln\left(\frac{N(T_0)}{\theta}\right) + T_0.$$

Let $T_1 = t_{c2} + T$, then the protein number at $t = T_1$ is $N(T_1) = N(T_0)e^{-r(T_1-T_0)}$. Suppose at time t_{c3}, $N(t_{c3}) = \theta$ again and:

$$\frac{dN(t)}{dt} = S - rN(t), \quad t \in (T_1, t_{c3} + T]. \tag{4.4.5}$$

The solution of (4.4.5) is:

$$N(t) = \frac{S}{r} + \left(N(T_1) - \frac{S}{r}\right)e^{-r(t-T_1)}.$$

t_{c3} is then solved by:

$$N(t_{c3}) = \frac{S}{r} + \left(N(T_1) - \frac{S}{r}\right)e^{-r(t_{c3}-T_1)} \Longrightarrow \quad t_{c3} = \frac{1}{r}\ln\left(\frac{S - rN(T_1)}{S - \theta r}\right) + T_1.$$

Let $T_2 = t_{c3} + T$, then protein number at $t = T_2$ is:

$$N(T_2) = \frac{S}{r} + \left(N(T_1) - \frac{S}{r}\right)e^{-r(T_2-T_1)}.$$

Now we write down the solution of $N(T_0), N(T_1)$ and $N(T_2)$ in detail:

$$\begin{aligned} N(T_0) &= \frac{S}{r}\left(1 - e^{-rT_0}\right) = \frac{S}{r}\left(1 - e^{-r(t_{c1}+T)}\right) \\ &= \frac{S}{r}\left(1 - e^{-rT + \ln\left(1 - \frac{\theta r}{S}\right)}\right) = \frac{S}{r}\left(1 - e^{-rT}\left(1 - \frac{\theta r}{S}\right)\right), \end{aligned}$$

$$N(T_1) = N(T_0)e^{-r(T_1-T_0)} = N(T_0)e^{-r(t_{c2}+T-T_0)}$$

$$= N(T_0)e^{-r\left(\frac{1}{r}\ln\left(\frac{N(T_0)}{\theta}\right)+T_0+T-T_0\right)} = N(T_0)e^{-rT}\frac{\theta}{N(T_0)} = \theta e^{-rT},$$

$$\begin{aligned} N(T_2) &= \frac{S}{r} + \left(N(T_1) - \frac{S}{r}\right)e^{-r(T_2-T_1)} = \frac{S}{r} - \frac{S-rN(T_1)}{r}e^{-r(t_{c3}+T-T_1)} \\ &= \frac{S}{r} - \frac{S-rN(T_1)}{r}e^{-r\left(\frac{1}{r}\ln\left(\frac{S-rN(T_1)}{S-\theta r}\right)+T_1+T-T_1\right)} \\ &= \frac{S}{r} - \frac{S-rN(T_1)}{r}e^{-rT}\frac{S-\theta r}{S-rN(T_1)} = \frac{S}{r}\left(1-e^{-rT}\left(1-\frac{\theta r}{S}\right)\right). \end{aligned}$$

$N(T_0) = N(T_2)$ indicates that the period of the oscillation is $T_p = T_2 - T_0$:

$$\begin{aligned} T_p = T_2 - T_0 &= t_{c3} + T - T_0 = \frac{1}{r}\ln\left(\frac{S-rN(T_1)}{S-\theta r}\right) + T_1 + T - T_0 \\ &= \frac{1}{r}\ln\left(\frac{S-rN(T_1)}{S-\theta r}\right) + t_{c2} + T + T - T_0 \\ &= 2T + \frac{1}{r}\ln\left(\frac{S-rN(T_1)}{S-\theta r}\right) + \frac{1}{r}\ln\left(\frac{N(T_0)}{\theta}\right) + T_0 - T_0 \\ &= 2T + \frac{1}{r}\ln\left(e^{-2rT} + \frac{S^2\left(1-e^{-rT}\right)}{\theta r(S-\theta r)}\right). \end{aligned} \tag{4.4.6}$$

Moreover, $N(t)_{\max} = N(T_2), N(t)_{\min} = N(T_1)$, so the amplitude of the oscillation is $A = N(T_2) - N(T_1)$:

$$A = N(T_2) - N(T_1) = \frac{S}{r}\left(1-e^{-rT}\left(1-\frac{\theta r}{S}\right)\right) - \theta e^{-rT} = \frac{S}{r}\left(1-e^{-rT}\right). \tag{4.4.7}$$

Subsequently, we investigate how the parameters control the period and the amplitude. T_p (4.4.6) is a composite function, we focus on the following term first:

$$Y = e^{-2rT} + \frac{S^2\left(1-e^{-rT}\right)}{\theta r(S-\theta r)}. \tag{4.4.8}$$

Taking derivatives of (4.4.8) with respect to S, T, θ, r, respectively:

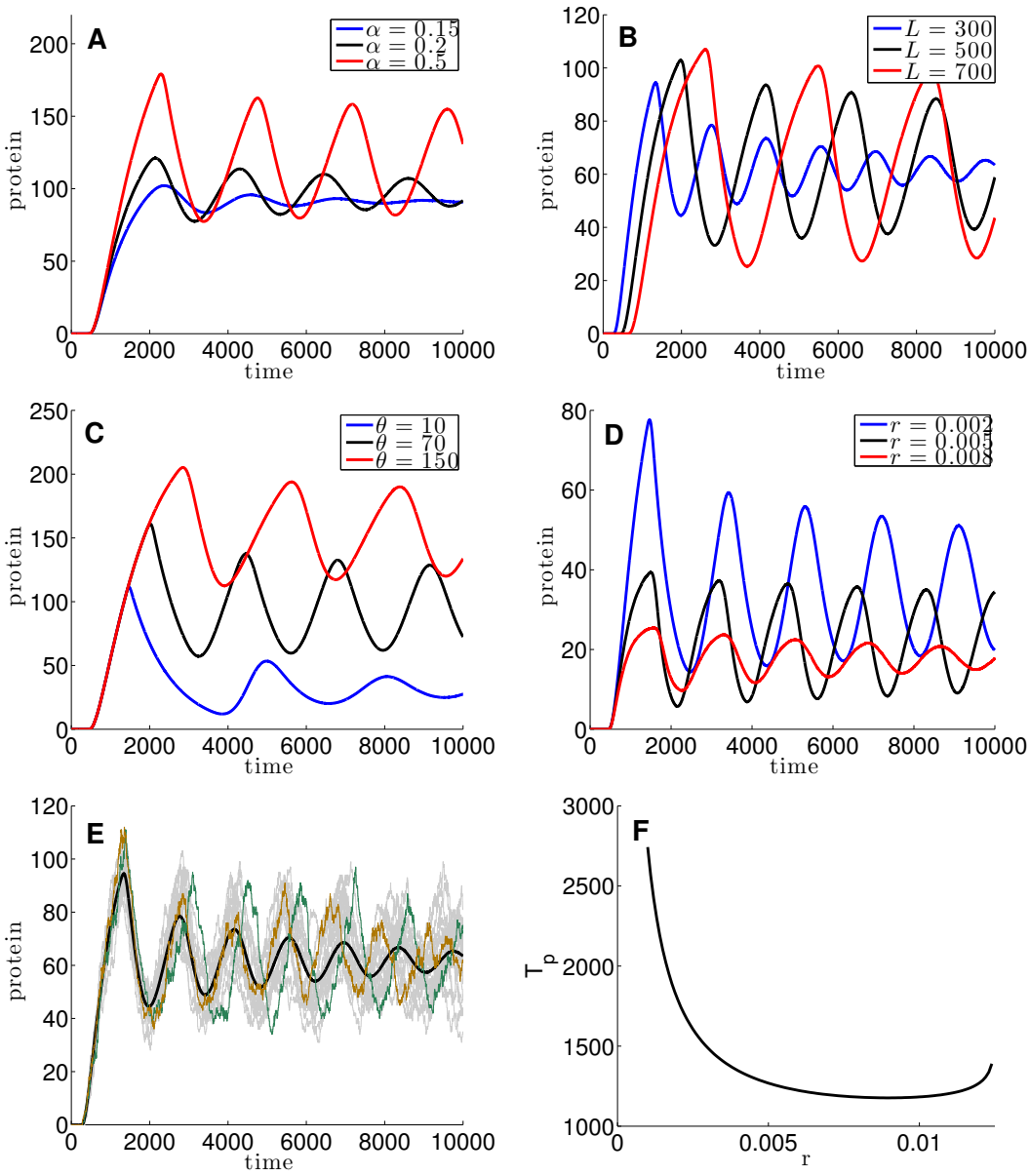


Figure 4.10: **The period and the amplitude of the oscillation for the simplified model.** Average of simulations is plotted. (A) Changing with α. $n = 10, L = 500, \theta = 100, r = 0.001, \beta = 0.7, \alpha = 0.15/0.2/0.5$. (B) Changing with L. $n = 5, \theta = 50, r = 0.002, \beta = 0.7, \alpha = 0.5, L = 300/500/700$. (C) Changing with θ. $n = 10, L = 500, r = 0.001, \beta = 0.7, \alpha = 0.5, \theta = 10/70/150$. (D) Changing with r. $n = 5, L = 500, \theta = 20, \beta = 0.7, \alpha = 0.3, r = 0.002/0.005/0.008$. (E) Average of simulations (black curve) is damping while single realisation of simulation (see green and brown) is oscillating. (F) Plotting (4.4.6) with respect to r. $\theta = 20, S = 0.25, T = 500$.

$$\frac{\partial Y}{\partial S} = \frac{S\left(1-e^{-rT}\right)(S-2\theta r)}{\theta r(S-\theta r)^2}, \quad > 0 \text{ if } S > 2\theta r, < 0 \text{ otherwise,}$$
$$\frac{\partial Y}{\partial T} = e^{-rT} r\left(\frac{S^2}{\theta r(S-\theta r)} - 2e^{-rT}\right) > e^{-rT} r\left(\frac{S^2}{S^2/4} - 2\right) = 2e^{-rT} r > 0,$$
$$\frac{\partial Y}{\partial \theta} = \frac{S^2\left(1-e^{-rT}\right)(2\theta r - S)}{\theta^2 r(S-\theta r)^2}, \quad > 0 \text{ if } \theta > \frac{S}{2r}, < 0 \text{ otherwise,}$$
$$\frac{\partial Y}{\partial r} = -2Te^{-2rT} + \frac{S^2\left(e^{-rT} rT(S-\theta r) - \left(1-e^{-rT}\right)(S-2\theta r)\right)}{\theta r^2(S-\theta r)^2},$$

however, $\frac{\partial Y}{\partial r}$ is hard to predict. Together with (4.4.6) and $S \propto \alpha(1-\alpha)$, $T \propto L$, we conclude that the period T_p decreases with α and θ first then increases with α and θ; increases with L. We need a numerical method to determine how T_p reacts to r. Taking derivatives of (4.4.7) with respect to S, T, θ, r, respectively:

$$\frac{\partial A}{\partial S} = \frac{1}{r}\left(1-e^{-rT}\right) > 0, \quad \frac{\partial A}{\partial T} = Se^{-rT} > 0, \quad \frac{\partial A}{\partial r} = \frac{Se^{-rT}}{r^2}\left(1+rT-e^{rT}\right) < 0.$$

We conclude that the amplitude A increases with α and L, decreases with r. A is independent of θ.

The above conclusion of the period and the amplitude of the limiting case $n \to \infty$ can be used to understand the period and the amplitude of the simplified model. In Figure 4.10, we show simulations to investigate the oscillations of the simplified model. From Figure 4.10A, we observe that the period decreases first and then increases with α and the amplitude increases with α. Figure 4.10B shows that a longer length of the lattice L leading to a larger period and a larger amplitude. Figure 4.10C illustrates that with the increasing of θ, the period is decreasing first and then increasing. From Figure 4.10D, we see that the period decreases with r first and then increases (just as Figure 4.10F shows), the amplitude decreases. Note that we observe that some averages of simulations are damped because of the shift of the oscillation phases due

to stochasticity, as Figure 4.10E shows. Though we have no quantitative prediction for the period and the amplitude, we do obtain qualitative prediction.

4.5 Conclusion

In this chapter, we studied a general case of feedback in which the translation process is repressed by proteins produced from that mRNA itself. Proteins first form protein complexes (we assumed protein complex formation is much faster than all the other biochemical reactions we considered in the model), then compete with ribosomes for binding to the ribosome binding site. The Monte Carlo method was used to obtain simulations. We found that in our model, protein level either stabilises or oscillates. To perform a mathematical analysis to obtain a basic understanding of the model, we simplified the model using some assumptions and approximations.

We assumed that protein complex binding/unbinding process is much faster than ribosome binding/unbinding to the ribosome binding site. $n\mathrm{P}\cdot\mathrm{RBS}$ was assumed to be in instantaneous equilibrium. Thus, the ribosome binding site was either occupied by a protein complex, or was free for ribosome to bind to. We applied equilibrium approximation on protein complex binding/unbinding process and obtained the probability that the ribosome binding site was free. We defined the effective ribosome binding site using that probability, and α_{eff} was given in (4.3.4). The simplified model was therefore a TASEP, whose ribosome binding rate was regulated by protein level. Mean field analysis was performed. The phase diagram of the simplified model is reshaped, the low density phase is extended at the expense of the high density phase and the maximal current phase. The impact factor, $\frac{1}{\theta r}$, is a key quantity in the simplified model. The impact factor measures the significance of the feedback. The larger the impact factor, the greater the effect the feedback has on the system. With the decreasing of

the impact factor, the simplified model degenerates to the standard TASEP eventually. In the high density phase and the maximal current phase, feedback does not affect the properties of the system. To be more precise, the system has the same properties as the TASEP without feedback. Protein level stabilises in these two phases. In the low density phase, we used a DDE model to gain a basic understanding of the basic dynamics. When $n > 1$ and condition (4.3.25) is satisfied, the system can have oscillatory behaviour.

To provide qualitative understanding of the period and the amplitude of the oscillation, we considered a limiting case of $n \to \infty$. The effective ribosome binding rate was approximated by the Heaviside function: $\alpha_{eff} = \alpha H(\theta - N)$. Oscillatory protein level can occur. The oscillation is due to the switching on/off of the translation when the protein level passes the threshold level, θ. A DDE model was developed to estimate the amplitude and the period of the oscillation. Results indicate that (i) with the increasing of α, the period decreases first and then increases, the amplitude increases; (ii) the period and the amplitude increase with L; (iii) with the increasing of θ, the period decreases first and then increases, the amplitude is independent of θ; (iv) with the increasing of r, the period decreases first and then increases, the amplitude decreases.

In the following chapter, we will consider a more complicated case of feedback. The translation process is repressed by the protein level on the one hand, and is promoted by ribosome recycling by forming a 'closed loop'.

Chapter 5

The Role of Auto-Negative Feedback and Ribosome Recycling in Regulating Translation

5.1 Introduction

As we mentioned in Chapter 1, the 3' end of an eukaryotic mRNA contains a structure called poly(A) tail, consisting of multiple adenosine monophosphates. Poly(A) binding protein (PABP) can bind to the poly(A) tail and further be bound by the 5' end of the mRNA, resulting in a 'closed loop' to enhance the efficiency of the translation process. In the 5' untranslated region (UTR) of a PABP mRNA, there commonly exists an A-rich sequence. Overexpression of PABP can down-regulate its own mRNA translation by binding to the A-rich sequence [81, 85, 91, 110, 144]. The autoregulation may be due to an interference with the translation initiation [4]. Essentially, Marshall *et al.* [104] assumed that ribosomes could be recruited directly from the stop codon to the start codon of the mRNA to initiate translation. In this chapter, we first give an

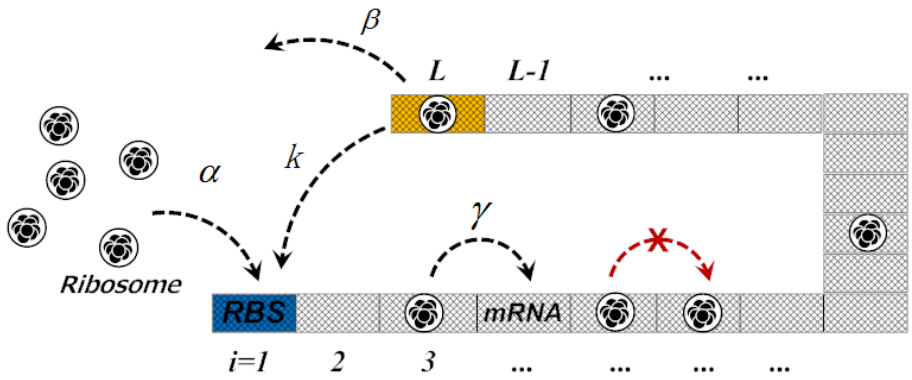


Figure 5.1: **Model of translation with ribosome recycling.** Ribosome on site L can move into site 1 at rate k to re-initiate translation process if $S_1 = 0$, or alternatively, it detaches from site L and enters into the particle reservoir. Please refer to the text for further detail. Schematic reproduced from [104].

introduction to the model of translation with ribosome recycling developed in [104]. By incorporating the autoregulation into this model, we develop a model of translation with ribosome recycling and auto-negative feedback.

5.2 Translation with Ribosome Recycling

In this section, we summarise the main results of the model in [104]. The schematic is shown in Figure 5.1. Ribosomes bind to site 1 at rate α if $S_1 = 0$, and jump from site i to site $i+1$ if $S_{i+1} = 0$. Ribosome on site L either detaches at rate β and enters the reservoir or moves into site 1 at the recycling rate k (if $S_1 = 0$) to re-initiate the translation process. Thus, in this model, the effective ribosome binding rate and effective ribosome exit rate can be derived as:

$$\alpha_{eff} = \alpha + k\rho_L, \qquad \beta_{eff} = \beta + k(1-\rho_1). \tag{5.2.1}$$

Applying the mean field framework of the TASEP, one can obtain the following results.

Low density phase. Substituting $\rho_1 = \alpha_{eff}$ and $\rho_L = \alpha_{eff}\left(1-\alpha_{eff}\right)/\beta_{eff}$ (see Lemma 2.5) into (5.2.1) yields:

$$\alpha_{eff} = \alpha + k\frac{\alpha_{eff}\left(1-\alpha_{eff}\right)}{\beta + k\left(1-\alpha_{eff}\right)} \quad \Longrightarrow \quad \alpha_{eff} = \frac{\alpha(\beta+k)}{\beta+\alpha k},$$

and

$$\beta_{eff} = \beta + k\left(1-\alpha_{eff}\right) = \beta + k\left(1 - \frac{\alpha(\beta+k)}{\beta+\alpha k}\right).$$

$J_{LD} = \alpha_{eff}\left(1-\alpha_{eff}\right)$ and $\rho_{LD} = \alpha_{eff}$ yields:

$$J_{LD} = \frac{\alpha\beta(\beta+k)(1-\alpha)}{(\beta+\alpha k)^2}, \qquad \rho_{LD} = \frac{\alpha(\beta+k)}{\beta+\alpha k}. \tag{5.2.2}$$

In the low density phase, $\alpha_{eff} < \beta_{eff}$ and $\alpha_{eff} < \frac{1}{2}$, substituting α_{eff} and β_{eff} into these inequalities yields:

$$\alpha < \begin{cases} \beta & \text{if} \quad 2\beta + k < 1, \\ \dfrac{\beta}{2\beta+k} & \text{otherwise.} \end{cases} \tag{5.2.3}$$

High density phase. Substituting $\rho_1 = 1-\beta_{eff}\left(1-\beta_{eff}\right)/\alpha_{eff}$ and $\rho_L = 1-\beta_{eff}$ (see Lemma 2.6) into (5.2.1) yields:

$$\beta_{eff} = \beta + k\left(1 - \left(1 - \frac{\beta_{eff}\left(1-\beta_{eff}\right)}{\alpha + k\left(1-\beta_{eff}\right)}\right)\right) \quad \Longrightarrow \quad \beta_{eff} = \frac{\beta(\alpha+k)}{\alpha+k\beta},$$

and

$$\alpha_{eff} = \alpha + k\left(1-\beta_{eff}\right) = \alpha + k\left(1 - \frac{\beta(\alpha+k)}{\alpha+k\beta}\right).$$

$J_{HD} = \beta_{eff}\left(1-\beta_{eff}\right)$ and $\rho_{HD} = 1-\beta_{eff}$ yields:

$$J_{HD} = \frac{\alpha\beta(\alpha+k)(1-\beta)}{(\alpha+k\beta)^2}, \qquad \rho_{HD} = \frac{\alpha(1-\beta)}{\alpha+k\beta}. \tag{5.2.4}$$

In the high density phase, $\beta_{eff} < \alpha_{eff}$ and $\beta_{eff} < \frac{1}{2}$, substituting α_{eff} and β_{eff} into these inequalities yields:

$$\alpha > \begin{cases} \beta & \text{if} \quad 2\alpha + k < 1, \\ \dfrac{k\beta}{1-2\beta} & \text{otherwise,} \end{cases} \qquad \beta < \frac{1}{2}. \tag{5.2.5}$$

Maximal current phase. Substituting $\rho_1 = 1 - \dfrac{1}{4\alpha_{eff}}$ and $\rho_L = \dfrac{1}{4\beta_{eff}}$ (see Lemma 2.8) into (5.2.1) yields:

$$\beta_{eff} = \beta + k\left(1 - \left(1 - \frac{1}{4\left(\alpha + k/4\beta_{eff}\right)}\right)\right) \quad \Longrightarrow \quad \beta_{eff} = \frac{\beta + \sqrt{\beta\left(\beta + \dfrac{k}{\alpha}\right)}}{2},$$

and

$$\alpha_{eff} = \alpha + k\frac{1}{4\beta_{eff}} = \alpha + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \dfrac{k}{\alpha}\right)}\right)}.$$

In the maximal current phase,

$$J_{MC} = \frac{1}{4}, \qquad \rho_{MC} = \frac{1}{2}. \tag{5.2.6}$$

Substituting α_{eff} and β_{eff} into the condition for the maximal current $\alpha_{eff}, \beta_{eff} \geqslant \frac{1}{2}$ yields:

$$\frac{\alpha}{2\alpha + k} \leqslant \begin{cases} \beta \leqslant \dfrac{k\alpha}{1-2\alpha} & \text{if} \quad \alpha + k < \frac{1}{2}, \\ \beta & \text{otherwise.} \end{cases} \tag{5.2.7}$$

The current and the average density for all the phases are therefore given by (5.2.2), (5.2.4) and (5.2.6). The boundaries are given by (5.2.3), (5.2.5) and (5.2.7). The reshaped phase diagram is shown in Figure 5.2A. The translation model with ribosome

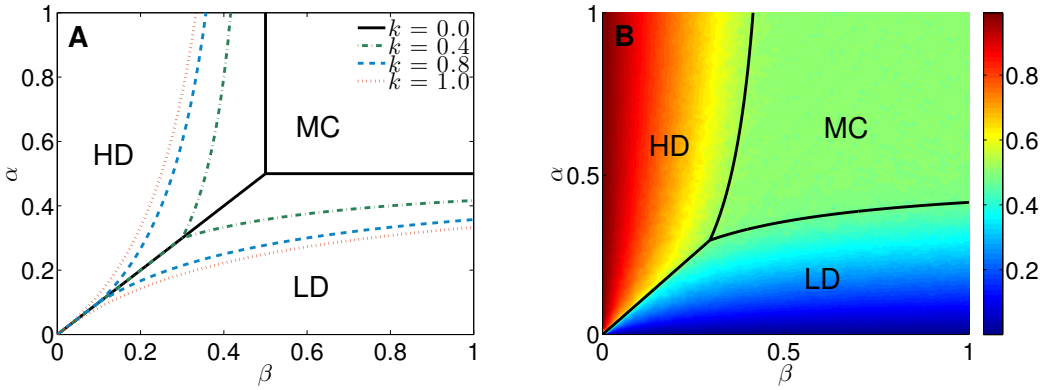


Figure 5.2: **Phase diagram for the translation model with ribosome recycling.** (A) Phase diagrams for different recycling rate k. (B) Average density simulated for $k = 0.4$, $L = 500$. Figure reproduced from [104].

recycling degenerates into the standard TASEP for $k = 0$. Moreover, the maximal current phase is extended at the expense of the low density phase and the high density phase. In this case, mRNA can access to optimal protein production rate for much lower α and β compared to the standard TASEP. An simulation sample is shown in Figure 5.2B, the simulation is in good agreement with the mean field solutions.

5.3 Model of Translation with Ribosome Recycling and Auto-Negative Feedback

Now we develop a model of translation with ribosome recycling and auto-negative feedback by extending the above model. In this model, there is a competition on the ribosome binding site of an mRNA between ribosomes and proteins produced from that mRNA itself. Following (4.3.4) defined for the model of translation with auto-negative feedback in Section 4, we now define:

$$\alpha_0 = \frac{\alpha}{1 + \left(\frac{N}{\theta}\right)^n}. \tag{5.3.1}$$

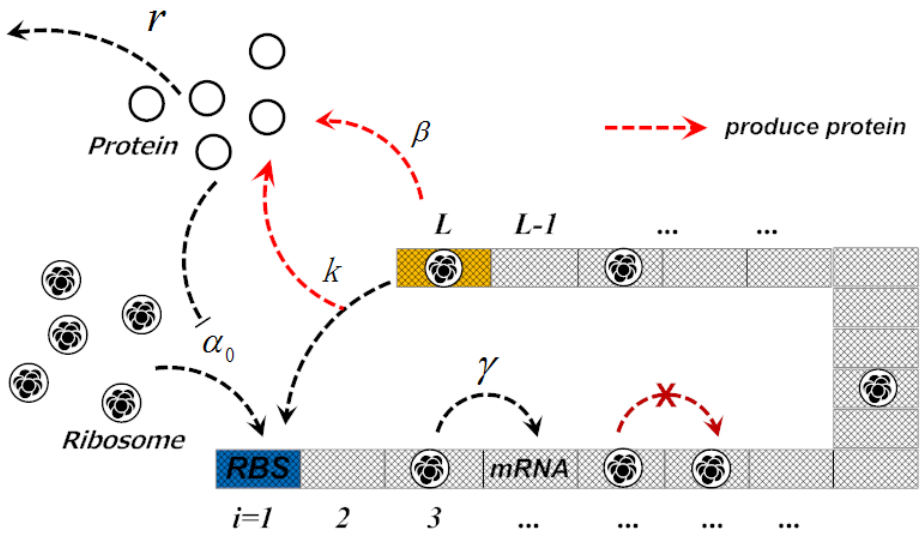


Figure 5.3: **Model of translation with ribosome recycling and auto-negative feedback.** Ribosome on site L moves into site 1 at rate k to re-initiate translation process if $S_1 = 0$, or alternatively, it detaches from site L and enters into the particle reservoir, both produce a protein (indicated by red dashed arrows). The ribosome binding rate is mediated by protein level, and is denoted by α_0. Proteins are removed at rate r.

Thus, ribosomes bind to site 1 at rate α_0, ribosome on site i moves to site $i+1$ at rate γ if site $i+1$ is vacant. Unless otherwise stated, $\gamma = 1$ in our model. Ribosome on site L moves into site 1 (if site 1 is vacant) at rate k and a protein is produced, or alternatively, moves into the ribosome reservoir at rate β and a protein is produced. Proteins are removed at rate r. From these set up, we are able to derive general expression for effective ribosome binding rate and effective ribosome exit rate:

$$\alpha_{eff} = \alpha_0 + k\rho_L = \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n} + k\rho_L, \qquad \beta_{eff} = \beta + k(1-\rho_1). \tag{5.3.2}$$

We use the Monte Carlo method to simulate the model. We construct another artificial site $L+1$, which is a copy of site L. If site $L+1$ is chosen, then ribosome recycling can take place. The algorithm is as follows:

1. Initialize TASEP: $S_i = 0$, $i \in [1, L]$, $N = 0$ and $t = 0$, set a MCS, τ,

2. Randomly choose a site:

   ①: If site 0 is chosen and $S_1 = 0$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \alpha_0 \tau$, set $S_1 = 1$;

   ②: If site i, $i \in [1, L-1]$ is chosen, and $S_i = 1$, $S_{i+1} = 0$, set $S_i = 0$, $S_{i+1} = 1$;

   ③: If site L is chosen and $S_L = 1$, draw a uniform $(0, 1)$ random number ξ: if $\xi < \beta \tau$, set $S_L = 0$;

   ④: If site $L+1$ is chosen and $S_L = 1$, $S_1 = 0$, draw a uniform $(0, 1)$ random number ξ: if $\xi < k\tau$, set $S_L = 0$ and $S_1 = 1$,

3. Repeat step 2 for $L+2$ times,

4. Set $c = 0$ (as a counter) and repeat the following procedure N times for protein removal: draw a uniform $(0, 1)$ random number ξ: if $\xi < r\tau$, set $c = c+1$,

5. Set $N = N - c$ and $t = t + \tau$, repeat step $2-3$ or else stop the process.

In following three subsections, we perform mean field analysis for all phases in turn. In the low density phase, we find a special case of oscillation. In the high density phase, we find a special case of bistable switching.

5.3.1 Maximal Current Phase

In the maximal current phase, we have the following mean field solutions:

$$\rho_1 = 1 - \frac{1}{4\alpha_{eff}}, \qquad \rho_L = \frac{1}{4\beta_{eff}}, \qquad N = \frac{1}{4r}. \tag{5.3.3}$$

Substituting (5.3.3) into (5.3.2) yields:

$$\beta_{eff} = \beta + k(1-\rho_1) = \beta + k\left(1 - \left(1 - \frac{1}{4\alpha_{eff}}\right)\right) = \beta + \frac{k}{4}\frac{1}{\alpha_{eff}}$$

$$= \beta + \frac{k}{4}\frac{1}{\frac{\alpha}{1+\left(1/(4\theta r)\right)^n} + \frac{k}{4\beta_{eff}}} = \beta + \frac{k}{\frac{4\alpha(4\theta r)^n}{(4\theta r)^n+1} + \frac{k}{\beta_{eff}}}$$
$$= \beta + \frac{k\beta_{eff}\left((4\theta r)^n+1\right)}{4^{n+1}\alpha(\theta r)^n\beta_{eff} + k\left((4\theta r)^n+1\right)},$$

multiply $4^{n+1}\alpha(\theta r)^n\beta_{eff} + k\left((4\theta r)^n+1\right)$ at the both sides:

$$\begin{aligned}&4^{n+1}\alpha(\theta r)^n\beta_{eff}^2 + k\beta_{eff}\left((4\theta r)^n+1\right)\\ =&\beta\left(4^{n+1}\alpha(\theta r)^n\beta_{eff} + k\left((4\theta r)^n+1\right)\right) + k\beta_{eff}\left((4\theta r)^n+1\right),\end{aligned}$$

therefore, we obtain a quadratic equation for β_{eff},

$$4^{n+1}\alpha(\theta r)^n\beta_{eff}^2 - 4^{n+1}\alpha\beta(\theta r)^n\beta_{eff} - k\beta\left((4\theta r)^n+1\right) = 0.$$

Subsequently,

$$\begin{aligned}\beta_{eff} &= \frac{4^{n+1}\alpha\beta(\theta r)^n \pm \sqrt{\left(4^{n+1}\alpha\beta(\theta r)^n\right)^2 + 4\cdot 4^{n+1}\alpha(\theta r)^n k\beta\left((4\theta r)^n+1\right)}}{2\cdot 4^{n+1}\alpha(\theta r)^n}\\ &= \frac{1}{2}\left(\beta \pm \sqrt{\beta^2 + \frac{k\beta\left((4\theta r)^n+1\right)}{4^n\alpha(\theta r)^n}}\right).\end{aligned}$$

Since we require β_{eff} to be positive and $\sqrt{\beta^2 + \frac{k\beta\left((4\theta r)^n+1\right)}{4^n\alpha(\theta r)^n}} > \beta$, therefore:

$$\beta_{eff} = \frac{1}{2}\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right). \tag{5.3.4}$$

Then α_{eff} is easy to obtain,

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n} + k\rho_L = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} + \frac{k}{4}\frac{1}{\beta_{eff}}$$

$$= \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right)}. \tag{5.3.5}$$

In the maximal current phase, the average density and the current are $\rho = \frac{1}{2}$ and $J = \frac{1}{4}$. Next we consider the phase diagram boundary for the maximal current phase. Knowing that in the maximal current phase, $\alpha_{eff}, \beta_{eff} \geqslant \frac{1}{2}$, we have:

$$\beta_{eff} = \frac{\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}}{2} \geqslant \frac{1}{2},$$

so $\sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)} \geqslant 1-\beta$. Since $1-\beta > 0$ for $\beta < 1$, we have

$$\beta^2 + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\beta \geqslant 1 - 2\beta + \beta^2,$$

$$\left(\frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right) + 2\right)\beta \geqslant 1,$$

therefore,

$$\beta \geqslant \frac{\alpha}{2\alpha + k\left(1+\left(\frac{1}{4\theta r}\right)^n\right)}, \tag{5.3.6}$$

and

$$\alpha_{eff} = \frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^n} + \frac{k}{2\left(\beta + \sqrt{\beta\left(\beta + \frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)\right)}\right)} \geqslant \frac{1}{2},$$

namely,

$$\frac{k}{2\left(\beta+\sqrt{\beta\left(\beta+\frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right)\right)}\right)} \geqslant \frac{1}{2}-\frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}}. \tag{5.3.7}$$

If $\frac{1}{2}-\frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}} \leqslant 0$, i.e., $\alpha \geqslant \frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right)$, then (5.3.7) is always true.

If $\alpha < \frac{1}{2}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right)$, we have:

$$k \geqslant 2\left(\frac{1}{2}-\frac{\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}}\right)\left(\beta+\sqrt{\beta\left(\beta+\frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right)\right)}\right),$$

subsequently,

$$k-\left(1-\frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}}\right)\beta \geqslant \left(1-\frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}}\right)\sqrt{\beta\left(\beta+\frac{k}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right)\right)} > 0. \tag{5.3.8}$$

From (5.3.8), we have a weak condition, $k-\left(1-\frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^{n}}\right)\beta > 0$, which can be rewritten as:

$$\beta < \frac{k}{1+\left(\frac{1}{4\theta r}\right)^{n}-2\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^{n}\right). \tag{5.3.9}$$

Squaring the both sides of (5.3.8) yields:

$$k^2 - 2\left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\beta k + \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)^2 \beta^2$$
$$\geqslant \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)^2 \beta^2 + \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)^2 \frac{k\beta}{\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right),$$

therefore,

$$k^2 \geqslant \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\beta k \left(2 + \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\frac{1+\left(\frac{1}{4\theta r}\right)^n}{\alpha}\right)$$
$$= \left(1 - \frac{2\alpha}{1+\left(\frac{1}{4\theta r}\right)^n}\right)\beta k \left(2 + \frac{1+\left(\frac{1}{4\theta r}\right)^n}{\alpha} - 2\right) = \left(1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha\right)\frac{\beta k}{\alpha},$$

So

$$\beta \leqslant \frac{\alpha k}{1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha}. \tag{5.3.10}$$

We have two conditions for β, (5.3.9) and (5.3.10). Since

$$\frac{k}{1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha}\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \frac{\alpha k}{1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha}$$
$$= \frac{k}{1+\left(\frac{1}{4\theta r}\right)^n - 2\alpha}\left(\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \alpha\right) > 0 \qquad (\alpha < 1),$$

so if (5.3.10) is satisfied, (5.3.9) is automatically satisfied. We finally obtain the phase

diagram boundary for the maximal current phase as follows:

$$\frac{\alpha}{2\alpha+k\left(1+\left(\frac{1}{4\theta r}\right)^n\right)} \leqslant \begin{cases} \beta \quad \leqslant \dfrac{\alpha k}{1+\left(\frac{1}{4\theta r}\right)^n-2\alpha} & \text{if} \quad \alpha < \dfrac{1}{2}\left(1+\left(\dfrac{1}{4\theta r}\right)^n\right), \\ \beta & \text{otherwise.} \end{cases} \tag{5.3.11}$$

5.3.2 Low Density Phase

In the low density phase, we have the following mean field solutions:

$$\rho_1 = \alpha_{eff}, \qquad \rho_L = \frac{\alpha_{eff}(1-\alpha_{eff})}{\beta_{eff}}, \qquad N = \frac{\alpha_{eff}(1-\alpha_{eff})}{r}. \tag{5.3.12}$$

Substituting (5.3.12) into (5.3.2) yields:

$$\begin{aligned} \alpha_{eff} &= \frac{\alpha}{1+\left(\frac{N}{\theta}\right)^n}+k\rho_L = \frac{\alpha}{1+\left(\frac{\alpha_{eff}(1-\alpha_{eff})}{\theta r}\right)^n}+k\frac{\alpha_{eff}(1-\alpha_{eff})}{\beta_{eff}} \\ &= \frac{\alpha}{1+\left(\frac{\alpha_{eff}(1-\alpha_{eff})}{\theta r}\right)^n}+k\frac{\alpha_{eff}(1-\alpha_{eff})}{\beta+k(1-\alpha_{eff})}. \end{aligned}$$

Subtracting $k\dfrac{\alpha_{eff}(1-\alpha_{eff})}{\beta+k(1-\alpha_{eff})}$ from the both sides, we get:

$$\begin{aligned} \alpha_{eff}-k\frac{\alpha_{eff}(1-\alpha_{eff})}{\beta+k(1-\alpha_{eff})} &= \frac{\alpha}{1+\left(\frac{\alpha_{eff}(1-\alpha_{eff})}{\theta r}\right)^n}, \\ \frac{\beta\alpha_{eff}+k\alpha_{eff}(1-\alpha_{eff})-k\alpha_{eff}(1-\alpha_{eff})}{\beta+k(1-\alpha_{eff})} &= \frac{\alpha(\theta r)^n}{(\theta r)^n+(\alpha_{eff}(1-\alpha_{eff}))^n}, \\ \frac{\beta\alpha_{eff}}{\beta+k(1-\alpha_{eff})} &= \frac{\alpha(\theta r)^n}{(\theta r)^n+(\alpha_{eff}(1-\alpha_{eff}))^n}, \end{aligned}$$

multiplying both sides by $\big(\beta+k(1-\alpha_{eff})\big)\big((\theta r)^n+(\alpha_{eff}(1-\alpha_{eff}))^n\big)$, we get:

$$\beta\alpha_{eff}(\theta r)^n+\beta\alpha_{eff}^{n+1}(1-\alpha_{eff})^n=\alpha(\theta r)^n\beta+\alpha(\theta r)^n k(1-\alpha_{eff}),$$

finally we obtain:

$$\beta\alpha_{eff}^{n+1}(1-\alpha_{eff})^n+(\theta r)^n(\beta+\alpha k)\alpha_{eff}-\alpha(\theta r)^n(\beta+k)=0. \tag{5.3.13}$$

In the low density phase, α_{eff} is the solution of (5.3.13) and $\beta_{eff}=\beta+k\big(1-\alpha_{eff}\big)$. The average density and the current are $\rho=\alpha_{eff}$ and $J=\alpha_{eff}\big(1-\alpha_{eff}\big)$. Next we consider the conditions of the existence of α_{eff} and the phase boundary for the low density phase. In the low density phase, $\alpha_{eff}<\beta_{eff}$, $\alpha_{eff}<\dfrac{1}{2}$. From $\alpha_{eff}<\beta_{eff}$, we have:

$$\alpha_{eff}<\beta_{eff}=\beta+k(1-\alpha_{eff}),\quad\Longrightarrow\quad(1+k)\alpha_{eff}<\beta+k,$$

therefore

$$\alpha_{eff}<\frac{\beta+k}{1+k}. \tag{5.3.14}$$

On the other hand, $\alpha_{eff}<\dfrac{1}{2}$. Therefore $\alpha_{eff}<\min\left(\dfrac{1}{2},\dfrac{\beta+k}{1+k}\right)$.

If $\dfrac{\beta+k}{1+k}\geqslant\dfrac{1}{2}$, i.e., $2\beta+2k\geqslant 1+k$, $2\beta+k\geqslant 1$, then $\alpha_{eff}<\dfrac{1}{2}$. Define $f=\beta\alpha_{eff}^{n+1}(1-\alpha_{eff})^n+(\theta r)^n(\beta+\alpha k)\alpha_{eff}-\alpha(\theta r)^n(\beta+k)$, the derivative with respect to α_{eff} is

$$\begin{aligned}f'&=(n+1)\beta\alpha_{eff}^n(1-\alpha_{eff})^n-n\beta\alpha_{eff}^{n+1}(1-\alpha_{eff})^{n-1}+(\theta r)^n(\beta+\alpha k)\\&=\beta\alpha_{eff}^n(1-\alpha_{eff})^{n-1}\big((n+1)(1-\alpha_{eff})-n\alpha_{eff}\big)+(\theta r)^n(\beta+\alpha k)\\&=\beta\alpha_{eff}^n(1-\alpha_{eff})^{n-1}\big(1-\alpha_{eff}+n(1-2\alpha_{eff})\big)+(\theta r)^n(\beta+\alpha k),\end{aligned}$$

since both $1-\alpha_{eff}$ and $1-2\alpha_{eff}$ are positive in $\left(0,\dfrac{1}{2}\right)$, so $f'>0$ in $\left(0,\dfrac{1}{2}\right)$. f is

increasing in $\left(0, \frac{1}{2}\right)$. Moreover, $f(0) = -\alpha(\theta r)^n(\beta+k) < 0$. Therefore, the solution of (5.3.13) exists if and only if $f\left(\frac{1}{2}\right) > 0$ (in this case, the solution is unique):

$$\begin{aligned} f\left(\frac{1}{2}\right) &= \beta\left(\frac{1}{2}\right)^{n+1}\left(1-\frac{1}{2}\right)^n + (\theta r)^n(\beta+\alpha k)\frac{1}{2} - \alpha(\theta r)^n(\beta+k) \\ &= \frac{\beta}{2}\left(\frac{1}{4}\right)^n + (\theta r)^n(\beta+\alpha k)\frac{1}{2} - \alpha(\theta r)^n(\beta+k) \\ &= \frac{(\theta r)^n}{2}\left(\beta\left(\frac{1}{4\theta r}\right)^n + \beta + \alpha k - 2\alpha(\beta+k)\right) \\ &= \frac{(\theta r)^n}{2}\left(\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \alpha(2\beta+k)\right), \end{aligned}$$

$f\left(\frac{1}{2}\right) > 0$ requires $\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right) - \alpha(2\beta+k) > 0$, namely,

$$\alpha < \frac{\beta}{2\beta+k}\left(1+\left(\frac{1}{4\theta r}\right)^n\right). \tag{5.3.15}$$

If $\frac{\beta+k}{1+k} < \frac{1}{2}$, i.e., $2\beta+k<1$, then $\alpha_{eff} < \frac{\beta+k}{1+k}$. Since f is increasing in $\left(0, \frac{1}{2}\right)$ and $f(0) < 0$, so the solution of (5.3.13) exists if and only if $f\left(\frac{\beta+k}{1+k}\right) > 0$ (the solution is unique):

$$\begin{aligned} f\left(\frac{\beta+k}{1+k}\right) &= \beta\left(\frac{\beta+k}{1+k}\right)^{n+1}\left(1-\frac{\beta+k}{1+k}\right)^n + (\theta r)^n(\beta+\alpha k)\frac{\beta+k}{1+k} - \alpha(\theta r)^n(\beta+k) \\ &= \beta\left(\frac{\beta+k}{1+k}\right)^{n+1}\left(\frac{1-\beta}{1+k}\right)^n + (\theta r)^n(\beta+\alpha k)\frac{\beta+k}{1+k} - \alpha(\theta r)^n(\beta+k) \\ &= \frac{(\beta+k)(\theta r)^n}{1+k}\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n + \beta + \alpha k - \alpha(1+k)\right) \\ &= \frac{(\beta+k)(\theta r)^n}{1+k}\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n + \beta - \alpha\right) > 0, \end{aligned}$$

which gives

$$\alpha < \beta \left(1 + \left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\right). \tag{5.3.16}$$

Overall, the phase diagram boundary for the low density phase are:

$$\alpha < \begin{cases} \beta \left(1 + \left(\dfrac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\right) & \text{if} \quad 2\beta + k < 1, \\ \dfrac{\beta}{2\beta+k}\left(1 + \left(\dfrac{1}{4\theta r}\right)^n\right) & \text{otherwise.} \end{cases} \tag{5.3.17}$$

Special case of oscillation. Motivated by the oscillatory protein level in the low density phase for the model in Chapter 4, we now develop a delay differential equation model to investigate oscillation for the model of translation with ribosome recycling and auto-negative feedback. The effective ribosome binding rate is given by:

$$\alpha_{eff} = \frac{\alpha}{1 + \left(N/\theta\right)^n} + k\rho_L.$$

We use the mean field solution to approximate ρ_L, namely,

$$\alpha_{eff} = \frac{\alpha}{1 + \left(N/\theta\right)^n} + k\rho_L = \frac{\alpha}{1 + \left(N/\theta\right)^n} + k\frac{\alpha_{eff}(1-\alpha_{eff})}{\beta + k(1-\alpha_{eff})},$$

subsequently, α_{eff} is constructed to be a function of protein number, N:

$$\alpha_{eff} = \frac{\alpha(\beta+k)}{\alpha k + \beta\left(1 + \left(N/\theta\right)^n\right)}. \tag{5.3.18}$$

Following 4.3.16, we propose the following DDE model:

$$\frac{dN(t)}{dt} = \frac{\alpha(\beta+k)}{\alpha k + \beta\left(1 + \left(\dfrac{N(t-T)}{\theta}\right)^n\right)}\left(1 - \frac{\alpha(\beta+k)}{\alpha k + \beta\left(1 + \left(\dfrac{N(t-T)}{\theta}\right)^n\right)}\right) - rN(t). \tag{5.3.19}$$

The steady state of (5.3.19), N^*, is the solution of:

$$R(N) = \frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\right) - rN = 0. \quad (5.3.20)$$

We now check if N^* exists within $\left(0, \frac{1}{4r}\right)$. Taking derivative of $R(N)$ with respect to N:

$$\begin{aligned}
\frac{dR(N)}{dt} &= \left(1-\frac{2\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\right)\frac{d}{dN}\left(\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\right) - r \\
&= \left(1-\frac{2\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\right)\frac{-\alpha\beta(\beta+k)n\left(\frac{N}{\theta}\right)^n}{N\left(\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)\right)^2} - r \\
&= -B - r < 0,
\end{aligned}$$

where

$$B = \left(1-\frac{2\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)}\right)\frac{\alpha\beta(\beta+k)n\left(\frac{N}{\theta}\right)^n}{N\left(\alpha k+\beta\left(1+\left(\frac{N}{\theta}\right)^n\right)\right)^2} > 0. \quad (5.3.21)$$

Hence, $R(N)$ is decreasing with N. Moreover,

$$R(0) = \frac{\alpha(\beta+k)}{\alpha k+\beta}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta}\right) = \frac{\alpha\beta(\beta+k)(1-\alpha)}{(\alpha k+\beta)^2} > 0,$$

$$R\left(\frac{1}{4r}\right) < \frac{1}{4} - r \cdot \frac{1}{4r} = 0.$$

Therefore, there exists a unique N^*. To check the stability of N^*, we consider a small perturbation $|\delta(t)| \ll 1$. Substituting $N(t) = N^* + \delta(t)$ into (5.3.19), using Taylor expansion and further dropping the higher order term yields:

$$\frac{d\delta(t)}{dt} = -B\delta(t-T) - r\delta(t). \tag{5.3.22}$$

Suppose that $\delta(t)$ has a solution $\delta(t) = Ae^{\lambda t}$ and substitute this solution into (5.3.22), we get:

$$\lambda + r = -Be^{-\lambda T}. \tag{5.3.23}$$

Any real solution λ of (5.3.23) is negative, since $\lambda = -(r + Be^{-\lambda T}) < 0$. Suppose that when $T = T_c$, the purely imaginary eigenvalues arise, i.e., $\lambda_c = \pm i\omega$, Hopf bifurcation occurs. We suppose that ω is real and positive. Given that $e^{ix} = \cos x + i \sin x$, substituting $\lambda = \pm i\omega$ into (5.3.23) yields:

$$\pm i\omega + r = -Be^{\mp i\omega T_c} \quad \Longrightarrow \quad \begin{cases} r = -B\cos\omega T_c \\ \omega = B\sin\omega T_c \end{cases}. \tag{5.3.24}$$

From (5.3.24), it follows that $\omega^2 = B^2(\sin\omega T_c)^2 = B^2 - B^2(\cos\omega T_c)^2 = B^2 - r^2$. So $B^2 - r^2 > 0$ is required. Since $B^2 - r^2 = (B+r)(B-r)$, and $B,\ r > 0$, the sign of $B^2 - r^2$ is determined by $B - r$. Using (5.3.20), (5.3.21), we are able to ascertain that:

$$B - r$$

$$= \left(1 - \frac{2\alpha(\beta+k)}{\alpha k + \beta\left(1 + \left(\frac{N^*}{\theta}\right)^n\right)}\right) \frac{\alpha\beta(\beta+k)n\left(\frac{N^*}{\theta}\right)^n}{N\left(\alpha k + \beta\left(1 + \left(\frac{N^*}{\theta}\right)^n\right)\right)^2}$$

$$
\begin{aligned}
&- \frac{\alpha(\beta+k)}{N^*\left(\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n\right)\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n\right)}\right)\\
&=\frac{\alpha\beta(\beta+k)}{\left(\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n\right)\right)^3} g\left(\left(\frac{N^*}{\theta}\right)^n\right),
\end{aligned}
$$

where

$$
\begin{aligned}
g\left(\left(\frac{N^*}{\theta}\right)^n\right) =& \beta(n-1)\left(\frac{N^*}{\theta}\right)^{2n} - (\beta+\alpha k)(1-\alpha)\\
&+\left(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k\right)\left(\frac{N^*}{\theta}\right)^n. \qquad (5.3.25)
\end{aligned}
$$

Hence, the sign of $B-r$ is determined by $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$, which is a quadratic function of $\left(\frac{N^*}{\theta}\right)^n$. First, we notice that if $n=1$, $g\left(\left(\frac{N^*}{\theta}\right)^n\right) = \left(-(\alpha+1)\beta-2\alpha k\right)\frac{N^*}{\theta} - (\beta+\alpha k)(1-\alpha) < 0$. Hence Hopf bifurcation cannot occur. Therefore, one requirement for Hopf bifurcation is that $n>1$.

For $n>1$, the discriminant of $g\left(\left(\frac{N^*}{\theta}\right)^n\right)$ is given by:

$$
\Delta = \left(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k\right)^2 + 4\beta(n-1)(\beta+\alpha k)(1-\alpha) > 0, \quad (5.3.26)
$$

provided $n>1$ and $\alpha<1$.

Therefore, $g\left(\left(\frac{N^*}{\theta}\right)^n\right)=0$ has two distinct real roots:

$$
\left(\frac{N^*}{\theta}\right)^n_{\pm} = \frac{-\left(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k\right) \pm \sqrt{\Delta}}{2\beta(n-1)},
$$

where Δ is given in (4.3.23) and $\left(\frac{N^*}{\theta}\right)^n_-$ is negative.

Hence, it is easy to conclude that $g\left(\left(\frac{N^*}{\theta}\right)^n\right) > 0$ requires:

$$\left(\frac{N^*}{\theta}\right)^n > \left(\frac{N^*}{\theta}\right)^n_+ = \frac{-\big(\beta(\alpha+n-2\alpha n-2)-(n+1)\alpha k\big)+\sqrt{\Delta}}{2\beta(n-1)}. \tag{5.3.27}$$

From (5.3.20) and (5.3.27), it follows that:

$$\begin{aligned}
&\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\right)\\
>&\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n\right)}\right) > r\theta\sqrt[n]{\left(\frac{N^*}{\theta}\right)^n_+},
\end{aligned}$$

namely, a Hopf bifurcation can occur if:

$$\theta r < \frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\left(1-\frac{\alpha(\beta+k)}{\alpha k+\beta\left(1+\left(\frac{N^*}{\theta}\right)^n_+\right)}\right)\frac{1}{\sqrt[n]{\left(\frac{N^*}{\theta}\right)^n_+}}. \tag{5.3.28}$$

5.3.3 High Density Phase

In the high density phase, we have the following mean field solutions:

$$\rho_1 = 1-\frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}}, \qquad \rho_L = 1-\beta_{eff}, \qquad N = \frac{\beta_{eff}(1-\beta_{eff})}{r}. \tag{5.3.29}$$

Substituting (5.3.29) into (5.3.2) yields:

$$\beta_{eff} = \beta + k(1-\rho_1) = \beta + k\left(1 - \left(1 - \frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}}\right)\right)$$
$$= \beta + k\frac{\frac{\beta_{eff}(1-\beta_{eff})}{\alpha}}{1+\left(\frac{N}{\theta}\right)^n} + k\rho_L = \beta + \frac{\frac{k\beta_{eff}(1-\beta_{eff})}{\alpha}}{1+\left(\frac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k(1-\beta_{eff}),$$

Subtracting $\dfrac{\frac{k\beta_{eff}(1-\beta_{eff})}{\alpha}}{1+\left(\frac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k(1-\beta_{eff})$ from the both sides, we get:

$$\beta_{eff} - \frac{\frac{k\beta_{eff}(1-\beta_{eff})}{\alpha}}{1+\left(\frac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k(1-\beta_{eff}) = \beta,$$

$$\frac{\frac{\alpha\beta_{eff}}{1+\left(\frac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k\beta_{eff}(1-\beta_{eff}) - k\beta_{eff}(1-\beta_{eff})}{\frac{\alpha}{1+\left(\frac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^n} + k(1-\beta_{eff})} = \beta,$$

$$\frac{\frac{\alpha(\theta r)^n\beta_{eff}}{(\theta r)^n + \left(\beta_{eff}(1-\beta_{eff})\right)^n}}{\frac{\alpha(\theta r)^n}{(\theta r)^n + \left(\beta_{eff}(1-\beta_{eff})\right)^n} + k(1-\beta_{eff})} = \beta,$$

$$\frac{\alpha(\theta r)^n\beta_{eff}}{\alpha(\theta r)^n + k(1-\beta_{eff})\left((\theta r)^n + \left(\beta_{eff}(1-\beta_{eff})\right)^n\right)} = \beta,$$

multiplying both sides by $\alpha(\theta r)^n + k(1-\beta_{eff})\left((\theta r)^n + \left(\beta_{eff}(1-\beta_{eff})\right)^n\right)$, we get:

$$\alpha(\theta r)^n\beta_{eff} = \alpha\beta(\theta r)^n + k\beta(1-\beta_{eff})(\theta r)^n + k\beta\beta_{eff}^n\left(1-\beta_{eff}\right)^{n+1},$$

we finally obtain:

$$k\beta\beta_{eff}^{n}\left(1-\beta_{eff}\right)^{n+1}-(k\beta+\alpha)(\theta r)^{n}\beta_{eff}+\beta(\theta r)^{n}(\alpha+k)=0. \tag{5.3.30}$$

In the high density phase, β_{eff} is the solution of (5.3.30). The effective initiation rate is then given by $\alpha_{eff}=\dfrac{\alpha}{1+\left(\dfrac{\beta_{eff}(1-\beta_{eff})}{\theta r}\right)^{n}}+k\left(1-\beta_{eff}\right)$. $\rho=1-\beta_{eff}$ and $J=\beta_{eff}(1-\beta_{eff})$. Next we consider the conditions of the existence of β_{eff} and the phase boundary for the high density phase. In the high density phase, $\alpha_{eff}>\beta_{eff}$, $\beta_{eff}<\dfrac{1}{2}$. From $\beta_{eff}<\dfrac{1}{2}$, we have $\dfrac{1}{2}>\beta_{eff}=\beta+k(1-\rho_1)>\beta$, namely, $\beta<\dfrac{1}{2}$ in high density phase.

$$\beta_{eff}=\beta+k(1-\rho_1)=\beta+k\left(1-\left(1-\frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}}\right)\right)=\beta+k\frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}},$$

From $\alpha_{eff}>\beta_{eff}$, we have

$$\beta_{eff}=\beta+k\frac{\beta_{eff}(1-\beta_{eff})}{\alpha_{eff}}<\beta+k(1-\beta_{eff})\quad\Longrightarrow\quad(1+k)\beta_{eff}<\beta+k,$$

subsequently,

$$\beta_{eff}<\frac{\beta+k}{1+k}. \tag{5.3.31}$$

On the other hand, $\beta_{eff}<\dfrac{1}{2}$, therefore, $\beta_{eff}<\min\left(\dfrac{1}{2},\dfrac{\beta+k}{1+k}\right)$.

Define $g=k\beta\beta_{eff}^{n}\left(1-\beta_{eff}\right)^{n+1}-(k\beta+\alpha)(\theta r)^{n}\beta_{eff}+\beta(\theta r)^{n}(\alpha+k)$, the derivative with respect to β_{eff} is:

$$\begin{aligned}g'&=k\beta\left(n\beta_{eff}^{n-1}\left(1-\beta_{eff}\right)^{n+1}-(n+1)\beta_{eff}^{n}\left(1-\beta_{eff}\right)^{n}\right)-(k\beta+\alpha)(\theta r)^{n}\\&=k\beta\beta_{eff}^{n-1}\left(1-\beta_{eff}\right)^{n}\left(n-(2n+1)\beta_{eff}\right)-(k\beta+\alpha)(\theta r)^{n},\end{aligned}$$

and the second derivative is:

$$\begin{aligned} g'' &= \left(k\beta\beta_{eff}^{n-1}\left(1-\beta_{eff}\right)^n\right)'\left(n-(2n+1)\beta_{eff}\right)-(2n+1)k\beta\beta_{eff}^{n-1}\left(1-\beta_{eff}\right)^n \\ &= nk\beta\beta_{eff}^{n-2}\left(1-\beta_{eff}\right)^{n-1}\left(2(2n+1)\beta_{eff}^2-4n\beta_{eff}+n-1\right). \end{aligned}$$

If $\frac{\beta+k}{1+k} \geqslant \frac{1}{2}$, i.e., $2\beta+k \geqslant 1$, then $\beta_{eff} < \frac{1}{2}$. We need to check the existence of the solution of (5.3.30) in $\left(0, \frac{1}{2}\right)$. The sign of g'' is determined by the quadratic term $2(2n+1)\beta_{eff}^2-4n\beta_{eff}+n-1$ which is associated with two roots:

$$\beta_{eff}^{\pm} = \frac{n}{2n+1} \pm \frac{\sqrt{2n+2}}{4n+2}.$$

Moreover,

$$\beta_{eff}^{+} = \frac{n}{2n+1}+\frac{\sqrt{2n+2}}{4n+2} = \frac{2n+1}{4n+2}+\frac{\sqrt{2n+2}-1}{4n+2} = \frac{1}{2}+\frac{\sqrt{2n+2}-1}{4n+2} > \frac{1}{2},$$

and

$$\beta_{eff}^{-} < \frac{1}{2}-\frac{\sqrt{2n+2}}{4n+2} < \frac{1}{2}, \quad \beta_{eff}^{-} = \frac{2n-\sqrt{2n+2}}{4n+2} = \frac{n-1}{2n+\sqrt{2n+1}} \geqslant 0.$$

Thus, $g'' > 0$ (g' is increasing) in $\left(0, \beta_{eff}^{-}\right)$ and $g'' < 0$ (g' is decreasing) in $\left[\beta_{eff}^{-}, \frac{1}{2}\right)$. $g'\left(\beta_{eff}^{-}\right)$ is the maximum and

$$\begin{aligned} g'(0) &= -(k\beta+\alpha)(\theta r)^n < 0, \\ g'\left(\frac{1}{2}\right) &= -k\beta\frac{1}{4^n}-(k\beta+\alpha)(\theta r)^n < 0. \end{aligned}$$

If $g'\left(\beta_{eff}^{-}\right) < 0$, then $g' < 0$ (g is decreasing) in $\left(0, \frac{1}{2}\right)$. Besides, $g(0) = \beta(\theta r)^n(\alpha+k) > 0$. Thus, the solution of (5.3.30) $g = 0$ exists if and only if $g\left(\frac{1}{2}\right) < 0$, and the

solution is unique.

If $g'\left(\beta_{eff}^{-}\right) > 0$, then there are two solutions to $g' = 0$. Let β_{eff}^{1} and β_{eff}^{2} denote the solutions and $0 < \beta_{eff}^{1} < \beta_{eff}^{2} < \frac{1}{2}$. Thus, $g' < 0$ (g is decreasing) in $\left(0, \beta_{eff}^{1}\right) \cup \left(\beta_{eff}^{2}, \frac{1}{2}\right)$, $g' > 0$ (g is increasing) in $\left(\beta_{eff}^{1}, \beta_{eff}^{2}\right)$. Moreover, $g(0) = \beta(\theta r)^n(\alpha + k) > 0$, and $g\left(\beta_{eff}^{1}\right)$ is the local minimum. We have following cases.

(I) If $g\left(\beta_{eff}^{1}\right) > 0$, then $g > 0$ in $\left(0, \beta_{eff}^{2}\right)$ and g is decreasing in $\left(\beta_{eff}^{2}, \frac{1}{2}\right)$. The solution of (5.3.30) $g = 0$ exists if and only if $g\left(\frac{1}{2}\right) < 0$, and the solution is unique.

(II) If $g\left(\beta_{eff}^{1}\right) < 0$ and $g\left(\beta_{eff}^{2}\right) < 0$, then $g < 0$ in $\left(\beta_{eff}^{1}, \frac{1}{2}\right)$, there exists a unique solution to (5.3.30) $g = 0$ in $\left(0, \beta_{eff}^{1}\right)$. In this case, $g\left(\frac{1}{2}\right) < 0$.

(III) If $g\left(\beta_{eff}^{1}\right) < 0$ and $g\left(\beta_{eff}^{2}\right) > 0$, then (5.3.30) $g = 0$ has three solutions if $g\left(\frac{1}{2}\right) < 0$ or two solutions if $g\left(\frac{1}{2}\right) > 0$. If $g\left(\frac{1}{2}\right) > 0$, we have:

$$\begin{aligned} g\left(\frac{1}{2}\right) &= k\beta\left(\frac{1}{2}\right)^n\left(1-\frac{1}{2}\right)^{n+1} - (k\beta+\alpha)(\theta r)^n\left(\frac{1}{2}\right) + \beta(\theta r)^n(\alpha+k) \\ &= \frac{(\theta r)^n}{2}\left(k\beta\left(\frac{1}{4\theta r}\right)^n - \alpha + 2\beta\alpha + \beta k\right) > 0, \end{aligned}$$

which gives:

$$\beta > \frac{\alpha}{2\alpha + k\left(1 + \left(\frac{1}{4\theta r}\right)^n\right)}. \tag{5.3.32}$$

However, from (5.3.11), we find that (5.3.32) overlaps the maximal current phase. $\beta_{eff} < \frac{1}{2}$ and $\beta_{eff} \geqslant \frac{1}{2}$ cannot be satisfied simultaneously. Thus there arises a contradiction. Therefore, in this case, $g\left(\frac{1}{2}\right) < 0$ and there are three solutions to (5.3.30) $g = 0$.

Overall, for $2\beta+k \geqslant 1$, the existence of the solution of β_{eff} requires $g\left(\frac{1}{2}\right)<0$:

$$\begin{aligned} g\left(\frac{1}{2}\right) &= k\beta\left(\frac{1}{2}\right)^n\left(1-\frac{1}{2}\right)^{n+1}-(k\beta+\alpha)(\theta r)^n\left(\frac{1}{2}\right)+\beta(\theta r)^n(\alpha+k) \\ &= \frac{(\theta r)^n}{2}\left(k\beta\left(\frac{1}{4\theta r}\right)^n-\alpha+2\beta\alpha+\beta k\right)<0, \end{aligned}$$

which gives:

$$\alpha > \frac{k\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right)}{1-2\beta}. \tag{5.3.33}$$

Note that one or three solutions can be obtained.

If $\frac{\beta+k}{1+k}<\frac{1}{2}$, i.e., $2\beta+k<1$, then $\beta_{eff}<\frac{\beta+k}{1+k}$. Similar to the case for $2\beta+k\geqslant 1$, the existence of the solution of β_{eff} requires $g\left(\frac{\beta+k}{1+k}\right)<0$ and one or three solutions can be obtained:

$$\begin{aligned} & g\left(\frac{\beta+k}{1+k}\right) \\ =& k\beta\left(\frac{\beta+k}{1+k}\right)^n\left(1-\frac{\beta+k}{1+k}\right)^{n+1}-(k\beta+\alpha)(\theta r)^n\frac{\beta+k}{1+k}+\beta(\theta r)^n(\alpha+k) \\ =& k\beta\left(\frac{\beta+k}{1+k}\right)^n\left(\frac{1-\beta}{1+k}\right)^{n+1}-(k\beta+\alpha)(\theta r)^n\frac{\beta+k}{1+k}+\beta(\theta r)^n(\alpha+k) \\ =& (\theta r)^n\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\frac{k(1-\beta)}{1+k}-\frac{(\beta+k)(k\beta+\alpha)}{1+k}+\beta(\alpha+k)\right) \\ =& (\theta r)^n\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\frac{k(1-\beta)}{1+k}+\frac{\beta(\alpha+k)(1+k)-(\beta+k)(k\beta+\alpha)}{1+k}\right) \\ =& (\theta r)^n\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\frac{k(1-\beta)}{1+k}+\frac{\alpha\beta k-\alpha k+\beta k-k\beta^2}{1+k}\right) \\ =& (\theta r)^n\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\frac{k(1-\beta)}{1+k}+\frac{k\alpha(\beta-1)+k\beta(1-\beta)}{1+k}\right) \\ =& (\theta r)^n\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\frac{k(1-\beta)}{1+k}+\frac{k(\beta-\alpha)(1-\beta)}{1+k}\right) \end{aligned}$$

$$=\frac{(\theta r)^n k(1-\beta)}{1+k}\left(\beta\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n+\beta-\alpha\right)<0,$$

which gives

$$\alpha>\beta\left(1+\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\right), \tag{5.3.34}$$

Overall, the phase diagram boundaries for the high density phase are:

$$\alpha>\begin{cases}\beta\left(1+\left(\frac{(\beta+k)(1-\beta)}{\theta r(1+k)^2}\right)^n\right) & \text{if} \quad 2\beta+k<1, \\ & \beta<\frac{1}{2}. \\ \frac{k\beta\left(1+\left(\frac{1}{4\theta r}\right)^n\right)}{1-2\beta} & \text{otherwise,}\end{cases} \tag{5.3.35}$$

5.3.4 Simulation and Discussion

The three phase diagram boundaries are given by (5.3.11), (5.3.17) and (5.3.35). The phases meet at the point $\alpha=\frac{1-k}{2}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)$, $\beta=\frac{1-k}{2}$. For $\theta r<\frac{\sqrt[n]{1-k}}{4\sqrt[n]{1+k}}$, $\frac{1-k}{2}\left(1+\left(\frac{1}{4\theta r}\right)^n\right)>1$ thus the maximal current phase vanishes.

We show a typical phase diagram in Figure 5.4. We use $n=1$ for Figure 5.4. According to the analysis above, it is known that for $n=1$, there is no oscillation in the low density phase and the solution of β_{eff} in the high density phase is unique. For the simulation shown in Figure 5.4A, we fixed all the other parameters and varied k. As we can see, for $k=0$, we recover the phase diagram for the auto-negative feedback model developed in Section 4, as would be expected. As k increases, the maximal current phase extends at the expense of the low density and high density phase. This is because increasing the ribosome recycling rate, k, promotes both α_{eff} and β_{eff}. Hence, the system can achieve optimal protein production (maximal current phase) with much lower values of α and β compared to the system without ribosome recycling. In Figure 5.4B, we fixed

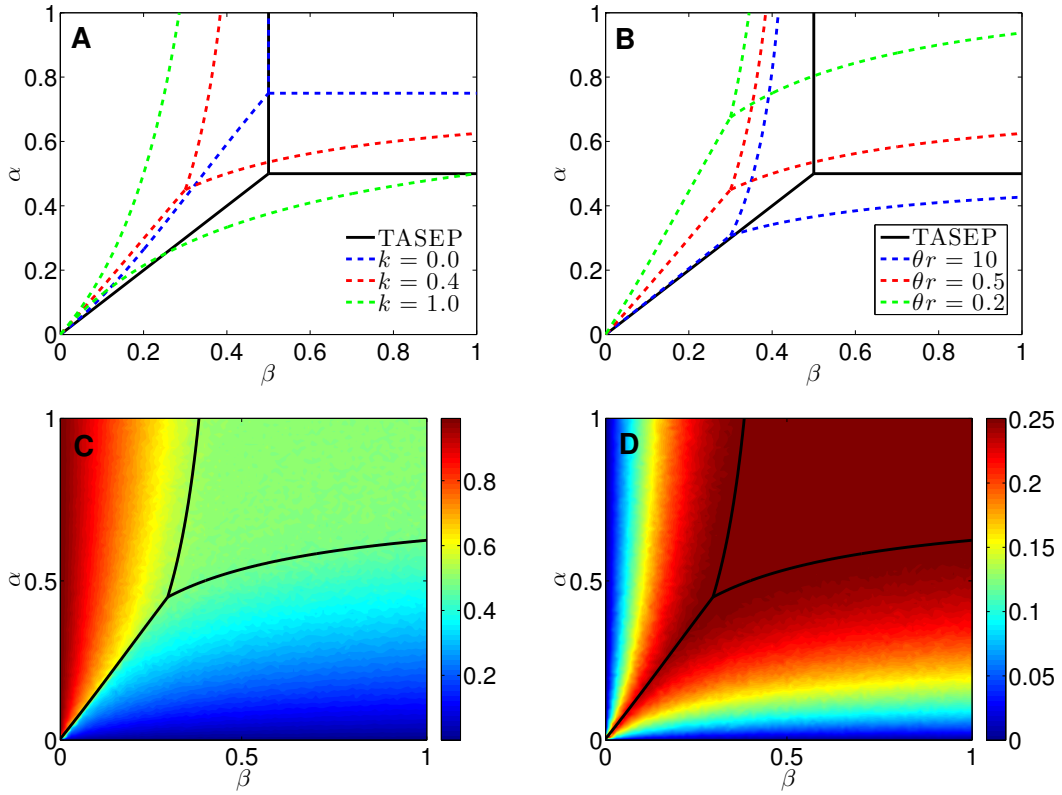


Figure 5.4: **Typical phase diagram for the translation model with ribosome recycling and auto-negative feedback.** (A) For $\theta r = 0.5$ and different values of k. (B) For $k = 0.4$ and different values of θr. (C) Average density and (D) current simulation for $\theta = 100$, $r = 0.005$, $k = 0.4$. Other parameters: $n = 1$, $L = 500$.

other parameters and varied θr. As we can see, for large θr ($\theta r = 10$), we recover the phase diagram for the model developed by Marshall *et al.* [104], as expected. As θr decreases, the low density phase extends at the expense of the high density and maximal current phases. This is because decreasing θr reduces α_{eff} directly and promotes β_{eff} indirectly. Hence, the system can achieve the low density phase with larger α and smaller β compared to the system without auto-negative feedback. Figure 5.4C and D show simulation results for the average density and current, revealing a good agreement between the mean field solution and simulation.

Next we consider the phase transitions. We are able to study how the average density and the current change with the parameters. Fixing β and varying α systematically,

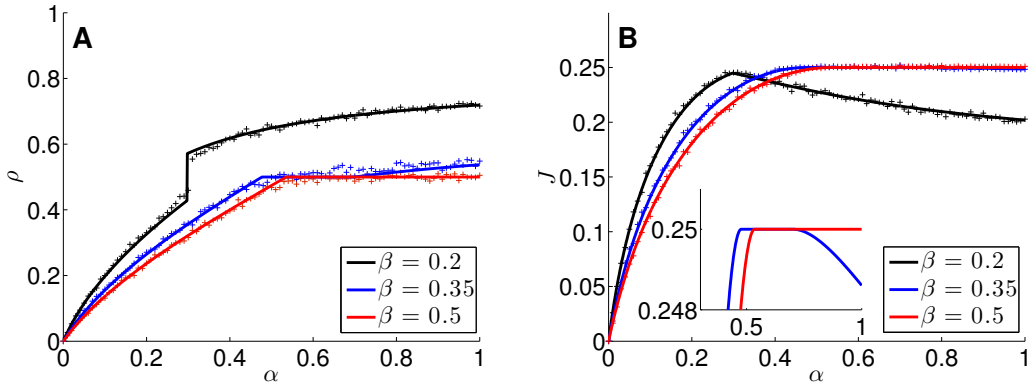


Figure 5.5: **Typical phase transitions for the translation model with ribosome recycling and auto-negative feedback.** (A) For the average density. (B) For the current. A zoom-in sub figure is provided. Solid lines are the mean field solutions, crosses are the simulations. $n = 1$, $k = 0.4$, $\theta = 100$, $r = 0.005$. $\beta = 0.2$ (balck, LD-HD), $\beta = 0.35$ (blue, LD-MC-HD), $\beta = 0.5$ (red, LD-MC).

the system can undergo a LD-HD transition, a LD-MC-HD transition and a LD-MC transition. Importantly, for $k = 1$, the boundary between the low density phase and the high density phase disappears, thus there is no LD-HD transition (see Figure 5.4A). In Figure 5.5, we show three possible phase transitions. (I) LD-HD transition (black in the figure). As α increases, so does the average density. The average density is discontinuous at a critical point, beyond which the system enters the high density phase from the low density phase. The current increases with α in the low density phase and decreases with α in the high density phase. The current attains the maximum at the critical point. That is due to the fact that in the low density phase, as α increases, α_{eff} increases, leading to the increase of the average density and the current (see (5.3.2) and (5.3.12)). Within the high density phase, α_{eff} keeps increasing with α, leading to an increase of ρ_1. Consequently, β_{eff} decreases. Thus, the average density increases and the current decreases (see (5.3.2) and (5.3.29)). (II) LD-MC-HD transition (blue in the figure). The average density increases with α in the low density and high density phases, but exhibits a plateau in the maximal current phase (Figure 5.5A). The current increases with α in the low density phase, exhibits a plateau in the maximal current phase, and decreases with α in the high density phase (Figure 5.5B). The plateau is

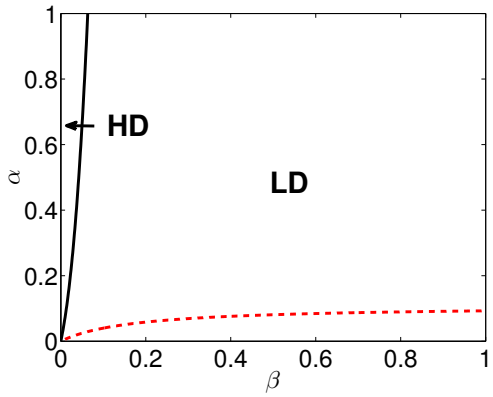


Figure 5.6: **Typical phase diagram for the translation model with ribosome recycling and auto-negative feedback: oscillation may occur.** Black line is the boundary between the low density phase and the high density phase. The maximal current phase vanishes. The red dashed line divides the low density phase into two regions. In the region above the red dashed line, oscillation can occur. The red dashed line is obtained by solving (5.3.28). $\theta = 50$, $r = 0.002$, $k = 0.2$, $n = 5$.

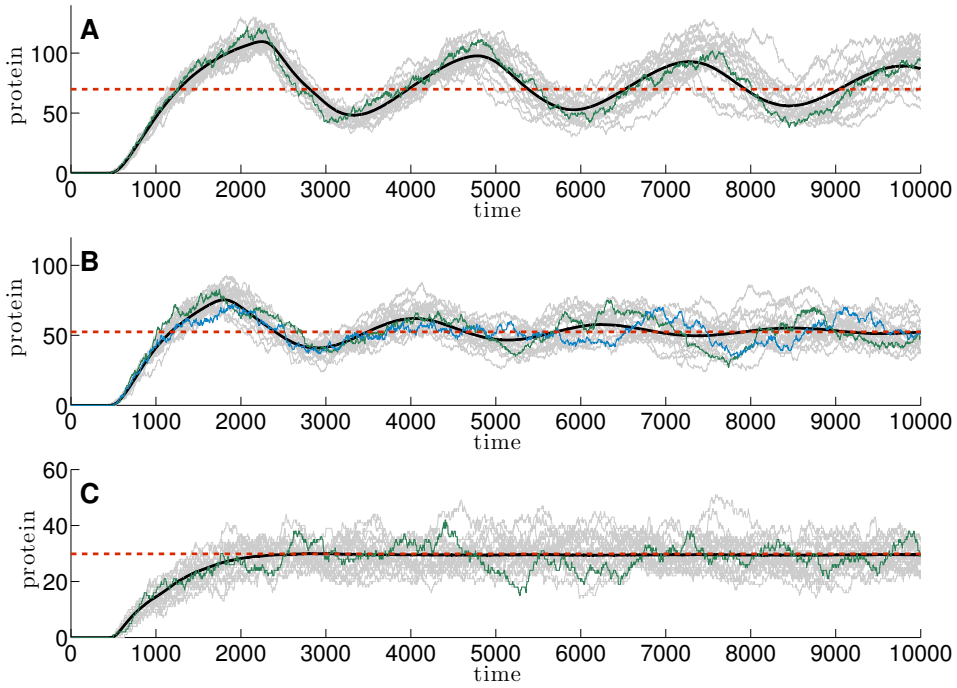


Figure 5.7: **Simulation for the low density phase of the translation model with ribosome recycling and auto-negative feedback: oscillation may occur.** Black lines are the averages of the 5000 realizations of stochastic simulations. Grey lines are 20 realizations of stochastic simulations, green (blue) line is one of them. Red dashed lines are the mean field solution $\alpha_{eff}\left(1-\alpha_{eff}\right)/r$, see (5.3.12). α_{eff} is obtained by solving (5.3.13) (A) $\alpha = 0.8$. (B) $\alpha = 0.2$. (C) $\alpha = 0.05$. $\beta = 0.5$, $\theta = 50$, $r = 0.002$, $k = 0.2$, $n = 5$, $L = 500$.

explained by the fact that in the maximal current phase, the average density is $\frac{1}{2}$ and the current is $\frac{1}{4}$. In the high density phase, α_{eff} keeps increasing with α, leading to an increase of ρ_1. Consequently, β_{eff} decreases and so does the current. (III) LD-MC transition (red in the figure). The average density and the current increases with α in the low density phase and exhibit a plateau in the maximal current phase.

It has been shown that in the low density phase, oscillatory behaviour can arise. We display a typical phase diagram with the existence of oscillation in Figure 5.6. The low density phase is divided into two regions by the condition (5.3.28). In the region above the red dashed line, oscillation can occur. We chose three sets of parameters to run the simulation, and present the results in Figure 5.7. Figure 5.7A/B exhibit oscillations ($\alpha = 0.8/0.2$, $\beta = 0.5$ is located in the oscillation region). However, we find that the average of simulation gives damped behaviour, rather than the standard oscillation. In 5.7B, we labelled two realisations of simulations with green and blue. It is observed that a single realisation of simulation is oscillating while the phases of these two realisations of simulation are shifted due to stochasticity, leading to a counteraction to the amplitude. Figure 5.7C exhibits stable protein level ($\alpha = 0.05$, $\beta = 0.5$ is located in the stability region).

It has been shown that in the high density phase, β_{eff} can have one or three solutions. We display a typical phase diagram with multiple solutions to β_{eff} in Figure 5.8A. In the region bounded by the red dashed line, β_{eff} has three possible solutions. We fixed β and varied α systematically to obtain the signal-response diagram Figure 5.8B/C. Figure 5.8B shows the LD-HD transition of the average density, ρ. Figure 5.8C indicates the LD-HD transition of the current, J. The signal-response diagrams give a higher stable steady state branch, a lower stable steady state branch and an unstable steady state branch, indicating that the system has three solutions within a range of

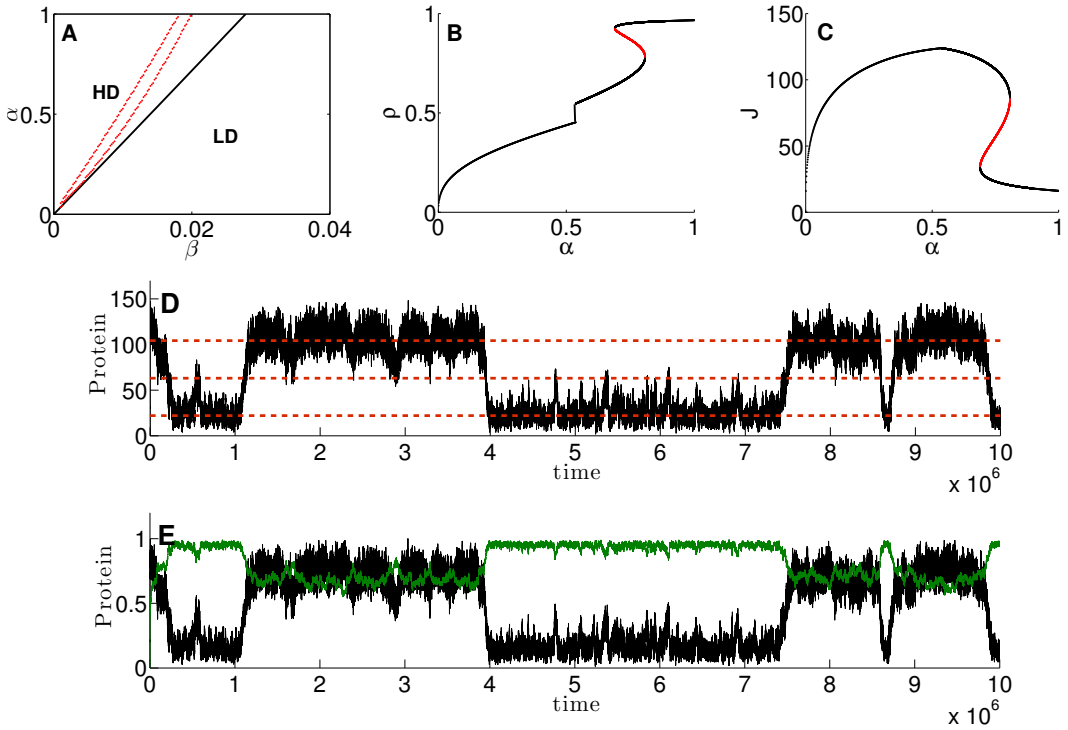


Figure 5.8: **Bistability in the high density phase of the translation model with ribosome recycling and auto-negative feedback.** (A) Phase diagram (partially shown). In the region bounded by the red lines, bistability occurs. The red lines are obtained by solving (5.3.30). Black line is the boundary between the low density phase and the high density phase. (B) Signal-response for ρ with respect to α. $\beta = 0.015$. (C) Signal-response for J with respect to α. $\beta = 0.015$. Red branches in (B) and (C) indicate unstable steady states. (D) Simulation for the protein level, red dashed lines are mean field solutions N, see (5.3.29), $\alpha = 0.77$, $\beta = 0.015$. (E) Normalised protein level (shown in (D)) and its corresponding average density $1-\beta_{eff}$ (shown as green). $\theta = 21$, $r = 0.002$, $k = 0.8$, $n = 2$, $L = 500$.

α. We notice that the current and the average density increase with α in the low density phase. As α increases, α_{eff} increases. In the low density phase, $\rho = \alpha_{eff}$ and $J = \alpha_{eff}\left(1-\alpha_{eff}\right)$, this is why ρ, J are increasing with α. In the high density phase, we observe a region where triple solutions of β_{eff} arise. In general, we find that as α increases, the average density increases and the current decreases. From (5.3.2), we see that α_{eff} keeps increasing with α, leading to an increase of ρ_1. Consequently, β_{eff} decreases. In the high density phase, $\rho = 1-\beta_{eff}$ and $J = \beta_{eff}\left(1-\beta_{eff}\right)$. This is the reason why ρ is increasing with α and J is decreasing with α. In Figure 5.8D,

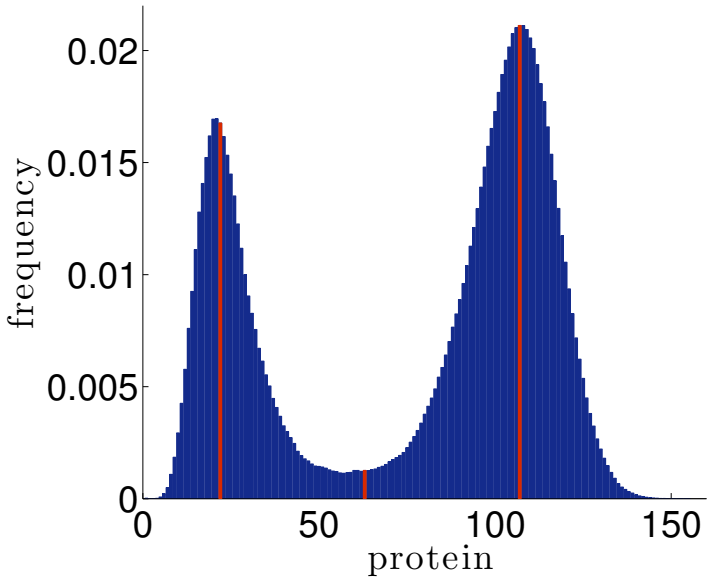


Figure 5.9: **Histogram for the bistable switch.** Run simulation for a long period of time with the same parameters for the protein level shown in Figure 5.8D and the corresponding histogram is plotted. Red bars are the frequencies for the steady states.

we present simulations for the protein level. The red dashed lines are the deterministic steady states. The upper and lower steady states are stable and the intermediate steady state is unstable. We can see that due to the stochasticity, the protein level shows a bistable switch between the two stable steady states. However, we can also observe that during the transition between the stable steady states, the protein level briefly stops at the unstable steady state before jumping to either the upper or the lower stable steady states. We run the simulation for a long period of time and obtain the corresponding histogram for the protein level shown in Figure 5.9. We can see that the system shows a bimodal distribution centred at the two stable steady states. In Figure 5.8E, we compare the normalised protein level with the corresponding average density. Higher average density gives a lower protein level. This can be explained as follows. Due to the stochasticity, the high protein level may drop to a certain level, which results in a significant increase of α_{eff}, and subsequently ρ_1. As a consequence, β_{eff} decreases, leading to an increase of the average density and a decrease of the current,

and thus protein level. Vice versa, the low protein level can rise to a certain level and thus α_{eff} and ρ_1 decreases significantly. As a consequence, β_{eff} increases, leading to a decrease of the average density and an increase of the current (and protein level).

5.4 Conclusion

In this chapter, we studied a translation model with ribosome recycling and auto-negative feedback that contains two novel features: (i) the effective ribosome binding/exit rate is enhanced by ribosome recycling, and (ii) the effective ribosome binding rate of a mRNA is negatively regulated by protein produced from itself. The phase diagram for the translation model with ribosome recycling and auto-negative feedback was derived by applying the mean field solution. Two quantities are important to the phase boundaries. One is the ribosome recycling rate k. k is responsible for increasing $\alpha_{eff}, \beta_{eff}$ to extend MC at the expense of LD and HD. The system can achieve MC with lower α, β compared to the standard TASEP. The other one is θr. θr is responsible for depressing α_{eff} and promoting β_{eff} to extend LD at the expense of HD and MC. In this model, the boundary between LD and HD can disappear for $k = 1$, MC can vanish for $\theta r < \frac{\sqrt[n]{1-k}}{4\sqrt[n]{1+k}}$. On altering α and β appropriately, the system undergoes LD-HD, LD-MC-HD, LD-MC transitions. In LD, the system can generate oscillatory behaviour. We developed a DDE model to investigate these oscillation and in particular, derived a condition for the existence of oscillations. In HD, the system can generate bistable behaviour. J, ρ, and the protein level can switch between the two stable states randomly. The oscillation and bistable switch behaviours are revealed in our model. However, at present we have no biological evidence for these behaviours. Thus, further experimental testing is required.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The translation process strongly affects cellular activities. Control of mRNA translation thus plays a pivotal role in the regulation of gene expression. In this thesis we presented a combination of mathematical modelling and numerical simulation to study and better understand the translation process. In order to do this we have taken a stochastic modelling approach and developed three models of translation with different regulatory factors. The first was a model of translation with mRNA degradation, for which it was assumed that the mRNA degradation was initiated by a simple single step. The second was a model of translation with auto-negative feedback, for which we assumed that the ribosome binding was negatively controlled by protein produced from its own mRNA. The last was a model of translation with ribosome recycling and auto-negative feedback, in which the ribosome at the termination stage was assumed to be able to access the ribosome binding site directly.

To provide a mathematical basis for the whole thesis, we gave an overview of the

fundamental model, the TASEP, in Chapter 2. The exact and mean field steady state solution of the TASEP were reproduced. Steps omitted in the original works [43, 44] were completed to provide the reader with a more comprehensive understanding of the results. We developed a linear approximation of transient state of the TASEP. The rate of particles entering the lattice was approximated as follows. Before the steady state of the entire lattice was reached, the rate of particles entering the lattice was determined by the particle motion in the leftmost part of the lattice. We defined the leftmost part of the lattice to be a sub-system with a particle binding rate, α, and a maximal exit rate, 1. Simulations suggested that the sub-system reached a steady state quickly and was in the low density phase if $\alpha < \frac{1}{2}$ and in the maximal current phase if $\alpha \geqslant \frac{1}{2}$. After the steady state of the entire lattice was reached, the rate of particles entering the lattice was given by the mean field solution of the current. The rate of the particles exiting the lattice was mainly approximated by two stages. The first stage was during $[0, L]$. In this stage, the rate of particles exiting the lattice was 0 because the very first ribosome bound to the lattice required an expected time L to reach site L from site 1 so that there was no particles leaving during $[0, L]$. The second stage was during $[L, t_c]$, where t_c was the time that the lattice reached the steady state. We assumed that the transient exit rate linearly increased in this stage. We obtained t_c using the argument that the net number of particles added to the lattice during transient state can be approximated by the total number of particles on the lattice in the steady state. We also provided a description of the Gillespie SSA and the Monte Carlo method in turn.

In Chapter 3, we studied a model of translation with mRNA degradation. We assumed that an mRNA in active translation is triggered to degrade by a simple single step: degradosome binding. Two types of mRNA degradation were considered: tracking degradation and rapid degradation. Our aim was to ascertain how the different degradation types affected protein production. We reported that the time that the degradosome binding event occurred was exponentially distributed. Subsequently, we

obtained the probability density functions for protein production from one mRNA during its lifetime. For the tracking degradation, the results are summarised as follows. In the low density phase, possible protein production is exponentially distributed. The average protein production can be promoted by increasing the ribosome binding rate or decreasing the degradosome binding rate. In the maximal current phase, possible protein production is also exponentially distributed. The average production is negatively related to the degradosome binding rate and is independent of all the other parameters. In the high density phase, we find that the protein production distribution is discontinuous at a critical point due to the change in the protein production rate. At the critical point, a peak is observed in the probability density function, indicating a higher likelihood of producing this critical amount of protein than that of its neighbours. The average protein production can be ramped up by increasing the ribosome exit rate and the length of the mRNA or decreasing the degradosome binding rate. Moreover, for $\alpha < \frac{1}{2}$, increasing ribosome binding rate also leads to a rise in production. On average, mRNA produces the most protein in the maximal current phase. If $\alpha < \frac{1}{2}$, mRNA produces more protein in the low density phase than in the high density phase. If $\alpha \geqslant \frac{1}{2}$, protein production in the low density phase and in the high density phase cannot be simply compared: determined by the value of β. For the rapid degradation, the most exceptional observation is that there is a high probability of an mRNA producing no proteins. This stems from the fact that mRNA can be rapidly degraded before the first ribosome on that mRNA successfully terminates the translation process. For all phases, decreasing the degradosome binding rate and the length of the mRNA lead to a rise in production. More particularly, in the low density phase, the average production is first increasing with the ribosome binding rate and then decreasing. In the high density phase, the average production increases with the ribosome exit rate. On average, mRNA produces the most protein in the maximal current phase. Comparison between the two mRNA degradation types shows that, given the same parameter configuration,

mRNA with the tracking degradation always produce more protein than that with the rapid degradation on average.

In Chapter 4, we studied a model of translation with auto-negative feedback control. The translation initiation was assumed to be repressed by proteins produced from that mRNA itself. We first considered the case where the negative feedback acted in a Hill function way. The effective ribosome binding rate was derived as a function of protein level. Applying the framework of the mean field solution of the TASEP, we were able to obtain the following results. In this case, feedback is measured by an impact factor, $\frac{1}{\theta r}$, where θ is the Hill function parameter shown in (4.3.4) and r is the protein removal rate. The larger the impact factor, the greater the effect the feedback has on the system. In the limit $\theta r \to \infty$, the system degenerates to the standard TASEP. The phase diagram of the model of translation with auto-negative feedback control is reshaped. The low density phase is extended at the expense of the high density phase and the maximal current phase. The maximal current phase vanishes for $\theta r < \frac{1}{4}$. In the high density phase and the maximal current phase, feedback does not affect the properties of the system. Namely, the same properties as the standard TASEP are obtained. The protein level reaches a steady state in these two phases. In the low density phase, due to the negative feedback and the intrinsic delay in the translation process, oscillatory behaviour can arise. We developed a DDE model to better understand these oscillations. Analysis showed that a Hopf bifurcation occurs when the Hill coefficient $n > 1$ and certain condition for θr (4.3.25) is satisfied. Using these conditions, we were able to divide the low density phase into two regions with or without oscillations. We next considered a limiting case of $n \to \infty$. The effective ribosome binding rate was derived as a Heaviside function. Oscillatory behaviour arises due to the switching on/off of the translation process when the protein level passes the threshold level, θ. A DDE model was developed to estimate the amplitude and the period of the oscillation. Qualitative prediction for the period and the amplitude were obtained. Results indicate that (i) with

the increasing of α, the period decreases first and then increases, while the amplitude increases; (ii) the period and the amplitude increase with L; (iii) with the increasing of θ, the period decreases first and then increases, while the amplitude is independent of θ; (iv) with the increasing of r, the period decreases first and then increases, while the amplitude decreases.

In Chapter 5, we studied a model of translation with ribosome recycling and auto-negative feedback. This model is an extension of the model in [104] and the model we studied in Chapter 4. The terminating ribosome was assumed to reinitiate the translation at a recycling rate k. In addition, the effective ribosome binding rate was negatively mediated by proteins produced from that mRNA itself. Effective ribosome binding/exit rates were derived that allowed us to apply the framework of the mean field solution of the TASEP. Two important quantities were found to reshape the phase diagram. One is k, which is responsible for enhancing the effective ribosome binding/exit rate to extend the maximal current phase at the expense of the low density phase and the high density phase. The other important quantity is the impact factor, $\frac{1}{\theta r}$, which is responsible for depressing the effective ribosome binding rate and enhancing the effective ribosome exit rate to extend the low density phase at the expense of the high density phase and the maximal current phase. The boundary between the low density phase and the high density phase disappears for $k = 1$ while the maximal current vanishes for $\theta r < \frac{\sqrt[n]{1-k}}{4\sqrt[n]{1+k}}$. The system undergoes LD-HD, LD-MC-HD, LD-MC transitions by fixing β and varying α systematically. In the maximal current phase, the properties of the system are the same as those in the TASEP. In the low density phase, due to the negative feedback and the intrinsic delay in the translation process, oscillatory behaviour can arise. Similar to the method followed in Chapter 4, a DDE model was developed to investigate the conditions for the existence of the Hopf bifurcation: the Hill coefficient $n > 1$ and a condition on θr (5.3.28). The low density phase is divided into two regions with/without oscillations. In the high density phase, due to the

nonlinear positive feedback, bistability arises. Stochastic simulation showed bistable switching between two stable steady states.

In this thesis, we have used mathematical models to study the translation process and its regulations. Each model focused on a particular regulation mechanism: mRNA stability, auto-negative feedback control, and coupling ribosome recycling and auto-negative feedback control. We have been able to better understand how protein is produced and how protein production is controlled. Novel theoretical behaviours, such as oscillation and bistable switching, have been revealed. However, no biological evidence is available to support these findings as of now.

6.2 Future Work

6.2.1 Synchrony in Multiple mRNA Translation

In this thesis, we have studied the translation process by considering a single mRNA using a TASEP method of investigation. However, in practice, multiple mRNAs undergo translation at the same time. One future direction would be to consider multiple mRNA translations. As we mentioned in Chapter 1, prokaryotic translation begins while the mRNA is still being synthesised whilst in eukaryotes, mRNA has to be transported to the cytoplasm before translation [59, 105]. Therefore, in either case, time lag between mRNA translations cannot be ignored. We begin the modelling of multiple mRNA translations by considering separate mRNAs with time lags. One may argue that in the prokaryotes, the translation process may be substantially affected by the coupling transcription process. However, it has been reported that in *E. coli*, the transcription of mRNA proceeds at a maximal speed of about $40-80$ bp/sec and that translation is carried out at a maximal speed of about 20 aa/sec [42, 148]. Since every

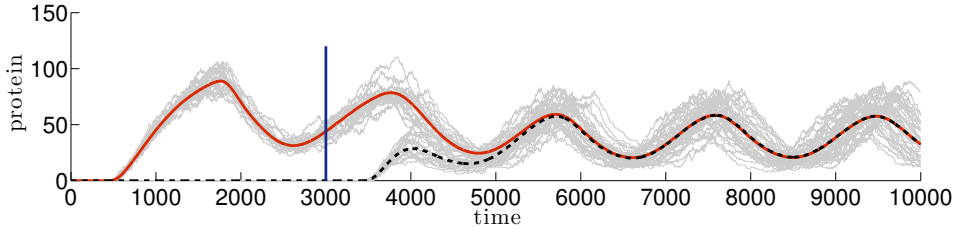


Figure 6.1: **Synchrony translation of mRNAs.** Red dashed/Black line is the average of 5000 stochastic simulations of the first/second mRNA. Grey lines are 20 realizations of simulations for each mRNA. Parameters: $L = 500$, $\alpha = 0.3$, $\beta = 0.7$, $\theta = 50$, $n = 5$, $r = 0.002$, $\Delta T = 3000$.

three base pairs code for one amino acid, the speeds of transcription and translation are nearly matched. Therefore we assume that the transcription process does not limit the movement of ribosome during translation. So in essence, we can ignore transcription altogether and just consider separate mRNAs. By introducing a time lag ΔT between the two identical mRNAs (using the model in Section 4.3 for each individual mRNA), we obtain the simulation shown in Figure 6.1. We find synchrony after some time. This synchronisation can be explained by the following simple system of differential equations that describe protein production:

$$\begin{cases} \dfrac{dN_1(t)}{dt} = f\left(N_{tot}(t-\Delta t)\right) - rN_1(t), \\ \dfrac{dN_2(t)}{dt} = H(t-\Delta T) f\left(N_{tot}(t-\Delta t)\right) - rN_2(t), \end{cases} \tag{6.2.1}$$

where $N_i(t)$ represents the number of protein from the ith mRNA, $N_{tot} = \sum N_i$, f is the common production rate function, ΔT is the time lag between two mRNA trsnaaltions, Δt is the intrinsic delay of translation and r is the protein removal rate. For $t > \Delta T$, subtracting the lower sub equation from the upper sub-equation of (6.2.1) yields:

$$\frac{d\Phi(t)}{dt} = -r\Phi(t), \tag{6.2.2}$$

where $\Phi(t) = N_1(t) - N_2(t)$. It follows immediately that $\Phi(t) = Ae^{-rt}$. Therefore $\Phi(t)$ tends to 0 as t tends to infinity, and synchrony arises.

The synchronisation provides an understanding of how a cell coordinates its function even without any further regulations. In the future, the synchrony model can be used to investigate how the protein production reacts by introducing more mRNAs into the system.

6.2.2 Another Application of TASEP: Ion Transport

The TASEP model method can be used to model many different phenomena besides mRNA translation, for example, to model ion transport. A possible future direction would be apply the TASEP method for potassium transport in the cell. The potassium channels typically consist of two components: a filter, which selects and allows potassium ions to pass, and a gate, which opens/closes the channel [49, 100]. The filter can accommodate potassium ions at 4 sites at the extracellular side. In addition, another two sites are designed for potassium ions to enter into or exit the membrane. The ion transport can be modelled using techniques such as TASEP [26, 80, 130]. In this section, we propose and study five alternative mechanisms for transport within transmembrane potassium channels based on the TASEP framework. Our working hypothesis is that movement of Potassium ions through the channel can be represented as a six-step process. Figure 6.2 provides a schematic representation of this assumption. We now provide further details of the assumptions behind this process.

1. Potassium ions pass through the membrane in a direction that takes them from high concentrations to low concentrations.

2. Only potassium ions can pass through channel.

3. The channel comprises four internal sites $S_1 - S_4$. There are two extra-channel sites S_0 and S_5: potassium ions start at S_0, pass through $S_1 - S_4$, enter into S_5 and finally leave the membrane channel. The state of each of these sites is a random

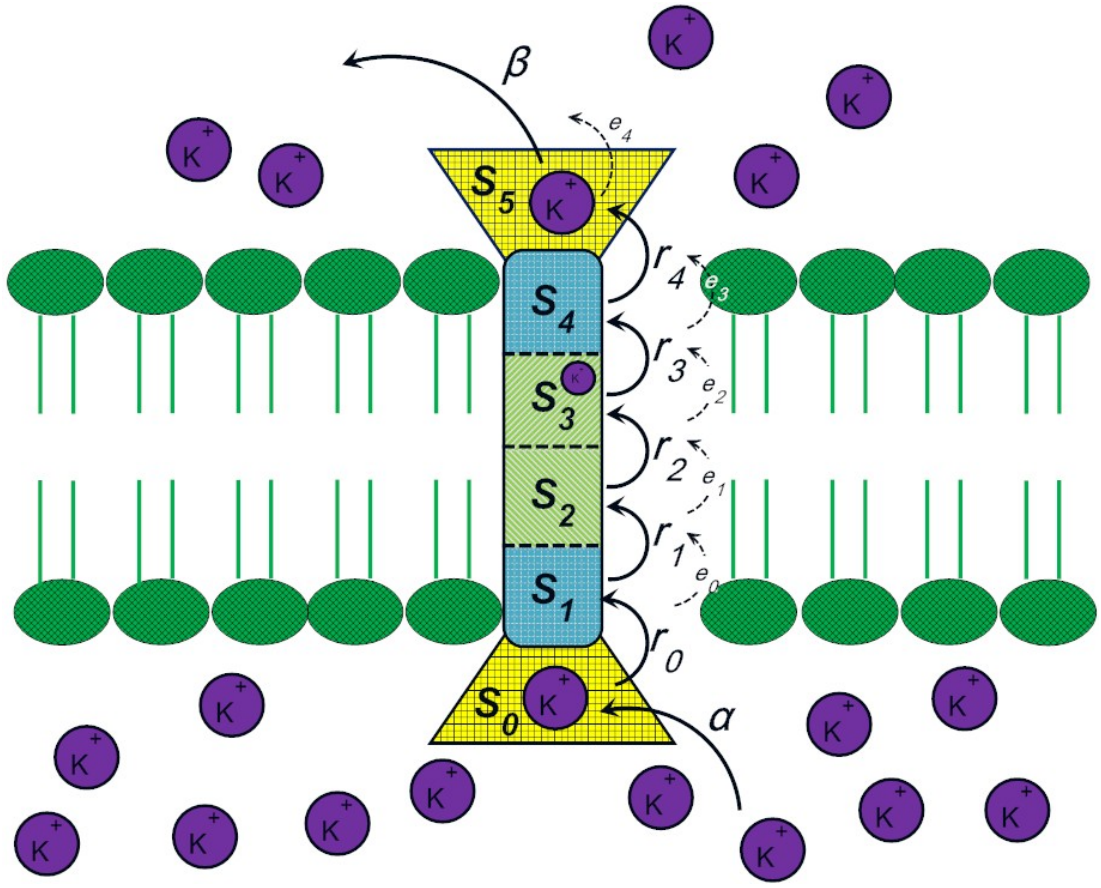


Figure 6.2: **Schematic of Potassium channel.** Potassium ions pass through the membrane from the high concentration side (intracellular) to low concentration side (extracellular). Potassium ions can bind onto the sites labelled by $S_1 - S_4$. Two extramembrane sites are represented S_0, S_5. Please refer to text for details.

variable: $S_i = 0$ (empty) or 1 (occupied), $i = 0..5$. Note that only one ion can occupy a site at any one time. Further experimental data suggests that S_2 and S_3 are always occupied.

4. The source of potassium ions within the cell is not limiting.

We now propose five models for how potassium ions pass through $S_0 - S_5$. We refer to these as (i) the TASEP, (ii) the Locust Model (iii) the Newton's Cradle, (iv) the Push Model and (v) the Water Model. The rules for each of these models are detailed as follows:

(i) **TASEP** We start with a simple model, a Totally Asymmetric Simple Exclusion Process or TASEP. In this model, potassium ions move from an infinite source to S_0 at rate α. Subsequently, each ion can move to next site at a fixed (but not necessarily equal) rate r_i, $i = 0,\ldots,4$ and finally leave S_5 at rate β. An ion can only move from position j to $j+1$ if $j+1$ is free.

(ii) **Locust Model** The basic rules are as for the TASEP model. It is further assumed that repulsion between ions enhances the hopping rate and thus additional rates e_i, $i = 0,\ldots,4$ are introduced. For instance, if $S_3 = 1$, $S_2 = 0$, the hopping rate of the ion in S_3 is r_3; if $S_3 = 1$, $S_2 = 1$, the the hopping rate of the ion in S_3 is then now $r_3 + e_2$. Hence, ions 'leap forward' as if pushed by (or jumping away from) the ion behind. The one-directional nature of this added 'force' is assumed to be a consequence of the potential difference across the channel.

(iii) **Newton's Cradle** In this model, potassium ions move from an infinite source to S_0 at rate α. If $S_1 = 0$, then this ion can move into S_1 at rate r_0. However, in addition, if $S_1 = \ldots = S_k = 1$, $k = 1,2,3,4$ or 5, then like Newton's Cradle, the ions in sites $S_0 - S_{k-1}$ remain still, but the ion in S_k moves to next site. Thus the ions in $S_1 - S_4$ cannot move to next site by themselves. However, the ion in S_5 can either be pushed or can simply leave at rate β.

(iv) **Push Model** In this model, ions can move into S_0 at rate α. If $S_1 = 0$, ion in S_0 can move to S_1 at rate r_0. If $S_1 = \ldots = S_k = 1$, $k = 1,2,3,4,5$, then all the ions in $S_0 - S_k$ move forward to next site at rate r_0 (pushed by the ion in S_0).

(v) **Water Model** It is assumed that the channel operates by the alternate transfer of water molecules and Potassium ions. Water molecules and Potassium ions competitively bind onto the S_0 site. The binding is successful if the first non-empty site is occupied by the other species, e.g. a K-ion binding event at S_0 is successful if a W-molecule occupies the first site. All subsequent movement

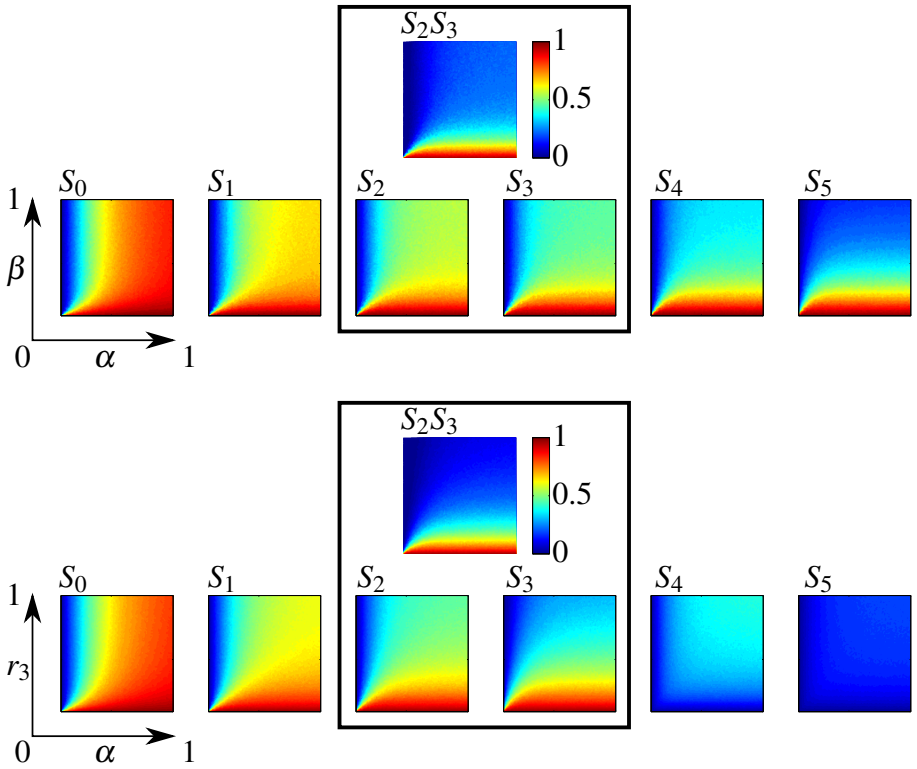


Figure 6.3: **Simulation results of the TASEP model.** The average densities of S_i, $i = 0,\ldots,5$ and S_2S_3 are displayed. Top sub figure is the results of the standard TASEP, $r_i = 0.5$, $i = 0,\ldots,4$, α, β are varying. Bottom is the results by varying α, r_3, $r_{0,1\,2,4} = 0.5$, $\beta = 1$.

rules follow the TASEP model.

Experimental data suggests that the sites S_2 and S_3 are occupied almost all the time. We focus on investigating high occupation of S_2S_3. The simulation for the TASEP model is shown in Figure 6.3. In the top sub figure, we vary α, β and fix $r_i = 0.5$, $i = 0,\ldots,4$. High densities of S_2, S_3, S_2S_3 are achieved when β is low. When β is low, ions queue on the channel so that the densities are high. In the bottom sub figure, we vary α, r_3, similar results are obtained. Small r_3 leads to the high average densities of S_2, S_3, even S_0, S_1. Small r_3 makes the site S_3 to be a slow site. The ions queue on $S_0 - S_3$.

The simulation of the Locust model is shown in Figure 6.4. All the average densities are low. Of course we can MAKE the densities of S_2, S_3 high by setting α to be very

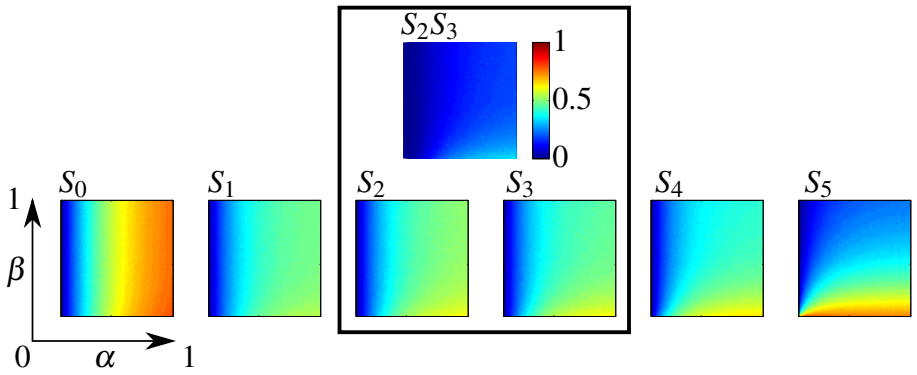


Figure 6.4: **Simulation result of the Locust model.** The average densities of S_i, $i = 0,\ldots,5$ and S_2S_3 are displayed. Fix $r_i = e_i = 0.5$, $i = 0,\ldots,4$, and vary α, β.

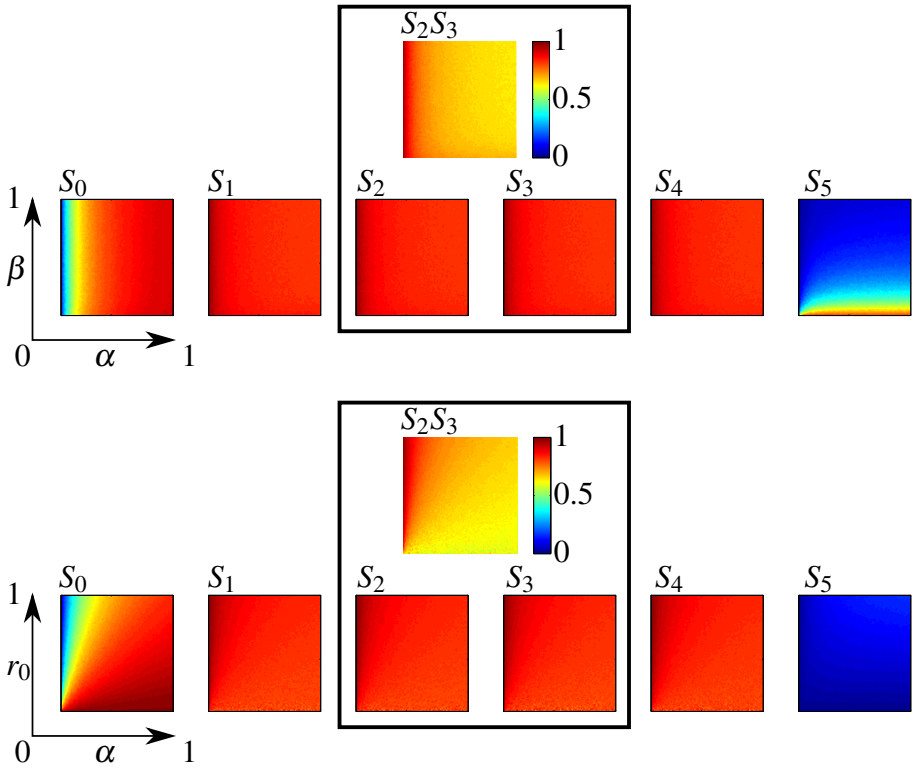


Figure 6.5: **Simulations of the Newton's Cradle model.** The average densities of S_i, $i = 0,\ldots,5$ and S_2S_3 are displayed. Top sub figure: $r_0 = 0.5$, α, β are varying. Bottom sub figure, $\beta = 1$, α, r_0 are varying

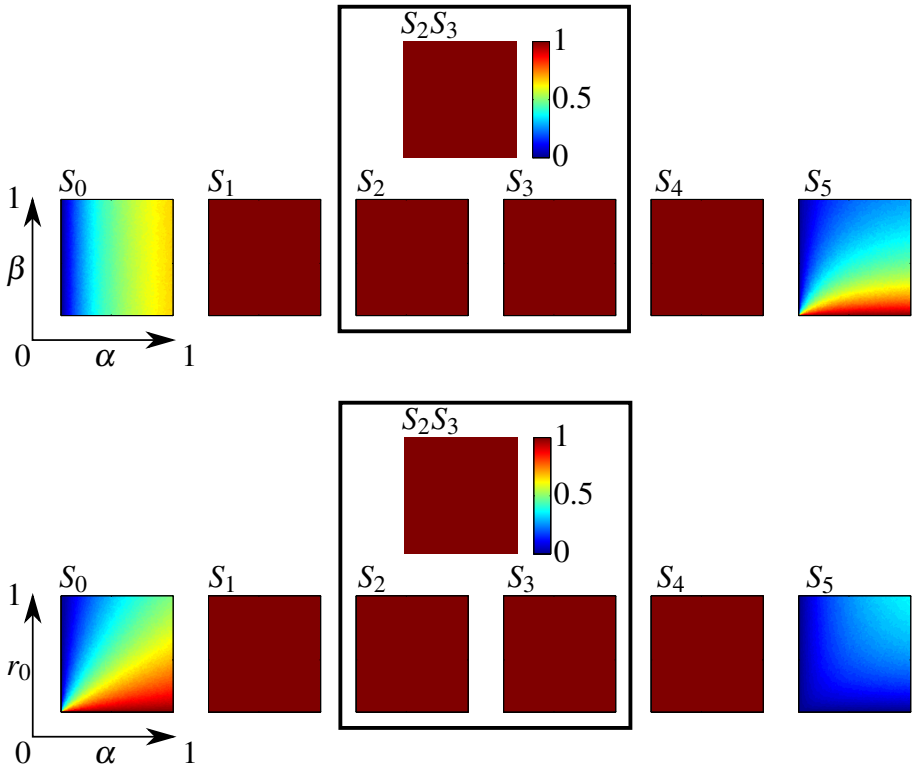


Figure 6.6: **Simulations of the Push model.** The average densities of S_i, $i = 0, \ldots, 5$ and S_2S_3 are displayed. Top sub figure: $r_0 = 0.5$, vary α, β. Bottom sub figure: $\beta = 1$, vary α, r_0.

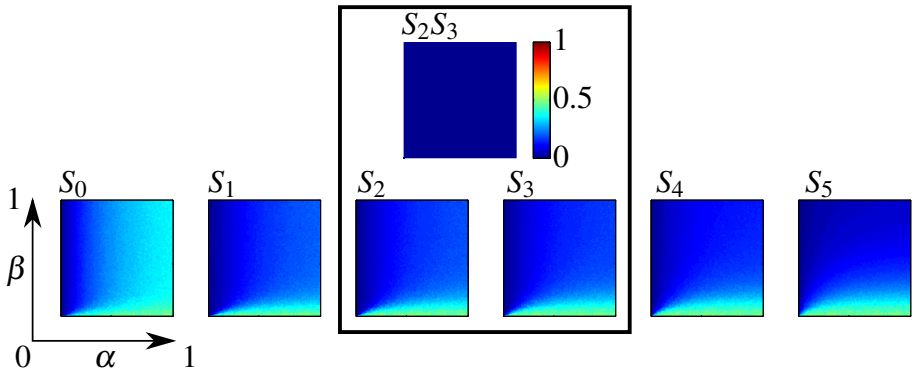


Figure 6.7: **Simulations of the Water model.** The average densities of S_i, $i = 0, \ldots, 5$ and S_2S_3 are displayed. Top sub figure: $r_0 = 0.5$, vary α, β. Bottom sub figure: $\beta = 1$, vary α, r_0.

large and r_3, e_2 to be very small, i.e., to make the site S_3 to be an artificial slow site. The simulation of the Newton's Cradle model is shown in Figure 6.5. In this model, only α, β, r_0 are involved. In the top sub figure, we fix $r_0 = 0.5$ and vary α, β. The average densities of S_2, S_3 are high, especially when α is small. Moreover, the average densities of S_2, S_3, S_2S_3 are independent on β. Thereafter, we generate the simulation by setting $\beta = 1$ and varying α, r_0. From the bottom sub figure, we conclude that small α and large r_0 lead to the high average densities of S_2, S_3, S_2S_3. The simulation of the Push model is shown in Figure 6.6. Similar to the Newton's model, only α, β, r_0 are involved. Either varying α, β or varying α, r_0 gives the same properties: the average densities of S_2, S_3, S_2S_3 are 1, i.e., they are always occupied. In the Water model, water and Potassium ion pass through the channel alternatively. Therefore, S_2 and S_3 can never be occupied by Potassium ions simultaneously. Just as Figure 6.7 shows.

We proposed five models. By choosing parameters carefully, we can achieve high densities of S_2, S_3, S_2S_3 for the TASEP model and the Locust model. Specifically, in the TASEP model, we can choose small r_3 (or r_4, β), so that the ions queue on the sites $S_0 - S_3$ (or $S_0 - S_4, S_0 - S_5$, respectively). In the Locust model, we can choose parameters to make a slow site. The slow site blocks the ions. For example, by setting small r_3, e_2, S_3 is made to be a slow site, ions queue on $S_0 - S_3$, such that the average densities of S_2, S_3 are high. In the Newton's Cradle model and the Push model, the average densities of S_2, S_3, S_2S_3 are dependent on α, r_0. To achieve high densities, the Newton's Cradle model requires small α and large r_0. In the Push model, the average densities of S_2, S_3, S_2S_3 are always 1. In the Water model, S_2 and S_3 can never be occupied by Potassium ions simultaneously. These models are developed to test the hypotheses of the potassium transport mechanisms, the sites S_2 and S_3 are occupied almost all the time. Though the four models (except the Water model) can all have high average densities of S_2, S_3, S_2S_3, we need more biological information to support these hypotheses.

Appendix A

Steady State: Mean Field Approximation vs Exact Solution

In this appendix, we will show comparison between exact solution and mean field approximation of TASEP for steady state.

The exact solution of TASEP is summarised as follows:

$$\langle S_i \rangle_{i<L} = \sum_{p=0}^{l-1} \frac{(2p)!}{p!(p+1)!} \frac{\langle W|C^{L-p-1}|V\rangle}{\langle W|C^L|V\rangle} + \frac{\langle W|C^{i-1}|V\rangle}{\langle W|C^L|V\rangle} \sum_{p=2}^{l+1} \frac{(p-1)(2l-p)!}{l!(l+1-p)!} \beta^{-p},$$

$$\langle S_L \rangle = \frac{1}{\beta} \frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle}, \qquad J = \frac{\langle W|C^{L-1}|V\rangle}{\langle W|C^L|V\rangle},$$

where $l = L - i$ and

$$\langle W|C^N|V\rangle = \sum_{p=1}^{N} \frac{p(2N-1-p)!}{N!(N-p)!} \frac{\left(1/\beta\right)^{p+1} - \left(1/\alpha\right)^{p+1}}{\left(1/\beta\right) - \left(1/\alpha\right)}.$$

The mean field approximation is summarised in following table:

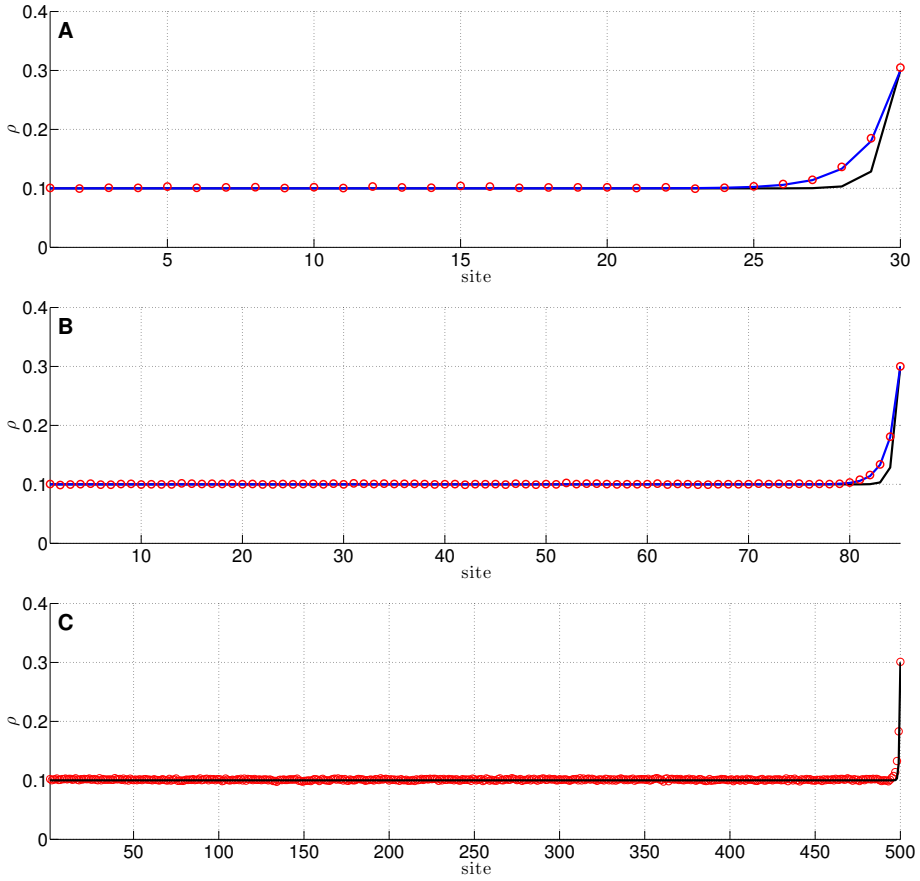


Figure A.1: **Comparison between exact solution and mean field approximation of TASEP.** Red circles are average density profiles over time from Monte Carlo simulation. Blue lines are exact steady state solution of TASEP shown in Lemma 2.4. Black lines are obtained by mean field recursion (2.2.5). (A) $L = 30$. (B) $L = 85$. (C) $L = 500$. $\alpha = 0.1, \beta = 0.3$.

Low density phase	$\alpha < \beta, \alpha < 1/2$	$\rho = \alpha$	$J = \alpha(1-\alpha)$
High density phase	$\alpha > \beta, \beta < 1/2$	$\rho = 1-\beta$	$J = \beta(1-\beta)$
Coexistence line phase	$\alpha = \beta < 1/2$	$\rho = 1/2$	$J = \alpha(1-\alpha)$
Maximal current phase	$\alpha, \beta \geqslant 1/2$	$\rho = 1/2$	$J = 1/4$

The comparison between exact solution and mean field approximation of TASEP is shown in Figure A.1. We can see that the mean field approximation fits the exact solution/simulation better with increasing of the size of TASEP L ($L = 30/85/500$ in the figure). As mentioned in Chapter 2, the mean field approximation is valid with sufficient large systems.

Appendix B

Rescale Real Time to Simulation Time

We rescale the time by the ribosome hopping rate γ. We take *E.coli* as the specific example. Liang *et.al* reported that the average hopping rate is 22 amino acids per second [96]. Therefore, the time unit of our simulation is $\frac{1}{22}$s. Let U be the time unit. About 80% of all mRNAs have half lives between 3 and 8 min [14]. We take the average 5.5 min. In passing, we comment on the relation between the mean life and half life in biological processes. We assume that the object we study is denoted by X and it decays exponentially with rate d:

$$\frac{dX(t)}{dt} = -dX(t) \quad \Longrightarrow \quad X(t) = X_0 e^{-dt}, \qquad X_0 = X(0).$$

Let T_m be mean life of X and T_h be half life. We have known that T_m is $1/d$. We are able to solve T_h by:

$$X(T_h) = X_0 e^{-dT_h} = \frac{1}{2} X_0,$$

and

$$T_h = (\ln 2)/d = T_m \ln 2.$$

Therefore the mean life time of mRNA is taken as $5.5/\ln 2$ min. To keep consistency of the unit:

$$\frac{5.5}{\ln 2}\,\text{min} = \frac{5.5}{\ln 2} \times 60\,\text{s} = \frac{5.5}{\ln 2} \times 60 \times 22\,\text{U} \approx 10^4\,\text{U}.$$

The average length of mRNA in prokaryotic cells is about $1000 - 1500$ nucleotides. Each codon consist of three nucleotides, so average length of prokaryotic mRNA is about $333 - 500$ codons. We use $L = 500$ in our simulations.

Appendix C

Monte Carlo Code for TASEP in C

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <time.h>
#define transient 100000
#define T 200000
#define L 500
int prob(double p)
{
  double r,pr;
  r=drand48();
  //printf("%lf\n",r);
  if(r<p)
    pr=1;
  else
    pr=0;
  return pr;
}
```

```
/*************************************************/
main(int argc, char *argv[])
{
  int i,j,k,site,jump,t;
  double alpha,beta,p,rho[L],mean_rho[(T-transient)],J,
              delta_alpha,mean_rho_av;
  int proteins,codon[L+1];
  FILE *f1_p,*f2_p;
  delta_alpha=0.1;
  beta=0.7;
  p=1.;
  srand(time(NULL));
  srand48(time(NULL));
  for(alpha=0.01;alpha<1.0;alpha=alpha+delta_alpha)
    {
      printf("alpha=%lf\n",alpha);
      proteins=0;
      for(i=0;i<(L+1);i++)
        codon[i]=0;
      for(i=0;i<L;i++)
          rho[i]=0.;
      for(t=0;t<T;t++)
        {
          for(i=0;i<=L;i++)
            {
              site=rand()%(L+1);
              if(site==0)
                {
                  jump=prob(alpha);
```

```
            if(jump==1 && codon[1]==0)
              codon[1]=1;
          }
        if(site==L && codon[site]==1)
          {
            jump=prob(beta);
            if(jump==1)
              {
                codon[site]=0;
                if(t>transient)
                  proteins++;
              }
          }
        if(site>0 && site<L && codon[site]==1)
          {
            jump=prob(p);
            if(jump==1 && codon[site+1]==0)
              {
                codon[site]=0;
                codon[site+1]=1;
              }
          }
      }
    if(t>transient)
      {
        for(k=0;k<L;k++)
          if(codon[k+1]==1)
            rho[k]++;
        mean_rho[t-transient]=0.;
```

```
                for(k=0;k<L;k++)
                  {
                    if(codon[k+1]==1)
                      mean_rho[t-transient]++;
                  }
                mean_rho[t-transient]=
                      mean_rho[t-transient]/(1.*L);
              }
          }
        mean_rho_av=0.;
        for(t=0;t<(T-transient);t++)
            mean_rho_av=mean_rho_av+mean_rho[t];
        mean_rho_av=mean_rho_av/(1.*(T-transient));
        J=(1.*proteins)/(1.*(T-transient));
        f1_p=fopen("alpha_J_rho_beta.dat","a");
        fprintf(f1_p,"%f\t%f\t%f\n",alpha,J,mean_rho_av);
        fclose(f1_p);
        f2_p=fopen("density_profile_beta.dat","a");
         for(k=0;k<L;k++)
           {
             rho[k]=rho[k]/(1.*(T-transient));
             fprintf(f2_p,"%lf\t%d\t%lf\n",alpha,k,rho[k]);
           }
         fprintf(f2_p,"\n");
         fclose(f2_p);
      }
  return 0;
}
```

Bibliography

[1] DA Adams, B Schmittmann, and RKP Zia. Far-from-equilibrium transport with constrained resources. *Journal of Statistical Mechanics-Theory and Experiment*, P06009, 2008.

[2] B Alberts, A Johnson, J Lewis, D Morgan, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell*, fourth edition. *Garland Science*, 2002.

[3] N Amrani, S Ghosh, DA Mangus, and A Jacobson. Translation factors promote the formation of two states of the closed-loop mRNP. *Nature*, 453(7199):1276–1285, 2008.

[4] J Bag. Feedback inhibition of poly(A)-binding protein mRNA translation - a possible mechanism of translation arrest by stalled 40S ribosomal subunits. *Journal of Biological Chemistry*, 276(50):47352–47360, 2001.

[5] S Bagga, J Bracht, S Hunter, K Massirer, J Holtz, R Eachus, and AE Pasquinelli. Regulation by *let-7* and *lin-4* miRNAs results in target mRNA degradation. *Cell*, 122(4):553–563, 2005.

[6] KE Baker and R Parker. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Current Opinion in Cell Biology*, 16(3):293–299, 2004.

[7] DJ Barnes and D Chu. *Introduction to Modeling for Biosciences*. *Springer*, pages 1–322, 2010.

[8] A Basu and D Chowdhury. Traffic of interacting ribosomes: effects of single-machine mechanochemistry on protein synthesis. *Physical Review E*, 75(2):021902, 2007.

[9] RT Batey. Structures of regulatory elements in mRNAs. *Current Opinion in Structural Biology*, 16(3):299–306, 2006.

[10] JG Belasco. All things must pass: contrasts and commonalities in eukaryotic and bacterial mRNA decay. *Nature Reviews Molecular Cell Biology*, 11(7):467–478, 2010.

[11] M Bengrine, A Benyoussef, H Ez-Zahraouy, and F Mhirech. Traffic model with quenched disorder. *Physical Letters A*, 253(3):135–138, 1999.

[12] R Benne and P Sloof. Evolution of the mitochondrial protein synthetic machinery. *Biosystems*, 21(1):51–68, 1987.

[13] JM Berg, JL Tymoczko, and L Stryer. *Biochemistry*, fifth dition. *W.H. Freeman*, 2002.

[14] JA Bernstein, AB Khodursky, PH Lin, S Lin-Chao, and SN Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceeding of the National Academy of Sciences of the United States of America*, 99(15):9697–9702, 2002.

[15] R Betney, E De Silva, J Krishnan, and I Stansfield. Autoregulatory systems controlling translation factor expression: Thermostat-like control of translational accuracy. *RNA-A Publication of the RNA Society*, 16(4):655–663, 2010.

[16] R Betney, E De Silva, C Mertens, Y Knox, J Krishnan, and I Stansfield. Regulation of release factor expression using a translational negative feedback loop: a systems analysis. *RNA-A Publication of the RNA Society*, 18(12):2320–2334, 2012.

[17] XP Bi and DJ Goss. Kinetic proofreading scanning models for eukaryotic translational initiation: the cap and poly(A) tail dependency of translation. *Journal of Theoretical Biology*, 207(2):145–157, 2000.

[18] WJ Blake, M Kaern, CR Cantor, and JJ Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–637, 2003.

[19] CA Brackley, MC Romano, C Grebogi, and M Thiel. Limited resources in a driven diffusion process. *Physical Review Letters*, 105(7):078102, 2010.

[20] C Brunel, P Romby, C Sacerdot, M Desmit, M Graffe, J Dondon, J Vanduin, B Ehresmann, C Ehresmann, and M Springer. Stabilised secondary structure at a ribosomal binding site enhances translational repression in *E. coli. Journal of Molecular Biology*, 253(2):277–290, 1995.

[21] AJ Carpousis. The *Escherichia coli* RNA degradosome: structure, function and relationship to other ribonucleolytic multienyzme complexes. *Biochemical Society Transactions*, 30(2):150–155, 2002.

[22] AJ Carpousis. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annual Review of Microbiology*, 61:71–87, 2007.

[23] LN Chen, RQ Wang, TJ Kobayashi, and K Aihara. Dynamics of gene regulatory networks with cell division cycle. *Phys. Rev. E*, 70(1):011909, 2004.

[24] T Chou. Ribosome recycling, diffusion and mRNA loop formation in translational regulation. *Biophysical Journal*, 85(2):755–773, 2003.

[25] T Chou and G Lakatos. Clustered bottlenecks in mRNA translation and protein synthesis. *Physical Review Letters*, 93(19):198101, 2004.

[26] T Chou and D Lohse. Entropy-driven pumping in zeolites and biological channels. *Physical Review Letters*, 82(17):3552–3555, 1999.

[27] L Ciandrini, I Stansfield, and MC Romano. Role of the particle's stepping cycle in an asymmetric exclusion process: a model of mRNA translation. *Physical Review E*, 81(5):051904, 2010.

[28] C Condon. RNA processing and degradation in *Bacillus subtilis*. *Microbiology and Molecular Biology Reviews*, 67(2):157, 2003.

[29] LJ Cook and RKP Zia. Feedback and fluctuations in a totally asymmetric simple exclusion process with finite resources. *Journal of Statistical Mechanics-Theory and Experiment*, P02012, 2009.

[30] LJ Cook and RKP Zia. Power spectra of a constrained totally asymmetric simple exclusion process. *Journal of Statistical Mechanics-Theory and Experiment*, P07014, 2010.

[31] LJ Cook, RKP Zia, and B Schmittmann. Competition between multiple totally asymmetric simple exclusion processes for a finite pool of resources. *Physical Review E*, 80(3):031142, 2009.

[32] F Crick. On protein synthesis. *The Symposia of the Society for Experimental Biology*, 12:138–163, 1958.

[33] F Crick. Central dogma of molecular biology. *Nature*, 227(5258):561, 1970.

[34] F Crick, L Barnett, S Brenner, and RJ Watts-Tobin. General nature of the genetic code for proteins. *Nature*, 192(4809):1227–1232, 1961.

[35] H De Jong. Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.

[36] V De Lorenzo and M Galperin. Microbial systems biology: bottom up and top down. *FEMS Microbiology Reviews*, 33(1):1–2, 2009.

[37] E De Silva, J Krishnan, R Betney, and I Stansfield. A mathematical modelling framework for elucidating the role of feedback control in translation termination. *Journal of Theoretical Biology*, 264(3):808–821, 2010.

[38] A Deana, Ha Celesnik, and JG Belasco. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature*, 451(7176):355–U14, 2008.

[39] J Dechargé and D Gogny. Hartree-Fock-Bogolyubov calculations with the D1 effective interaction on spherical nuclei. *Phys. Rev. C*, 21:1568–1593, 1980.

[40] F Delvillani, G Papiani, G Deho, and F Briani. S1 ribosomal protein and the interplay between translation and mRNA decay. *Nucleic Acids Research*, 39(17):7702–7715, 2011.

[41] C Deneke, S Rudorf, and A Valleriani. Transient phenomena in gene expression after induction of transcription. *PLoS ONE*, 7(4), 2012.

[42] PP Dennis and H Bremer. Differential rate of ribosomal protein synthesis in *Escherichia coli*. *Journal of Molecular Biology*, 84(3):407–422, 1974.

[43] B Derrida, E Domany, and D Mukamel. An exact solution of a one-dimensional asymmetric exclusion model with open boundaries. *Journal of Statistical Physics*, 69(3-4):667–687, 1992.

[44] B Derrida, MR Evans, V Hakim, and V Pasquier. Exact solution of a 1-D asymmetric exclusion model using a matrix formulation. *Journal of Physics A-Mathematical and General*, 26(7):1493–1517, 1993.

[45] RJ Dimelow and SJ Wilkinson. Control of translation initiation: a model-based analysis from limited experimental data. *Journal of the Royal Society Interface*, 6(30):51–61, 2009.

[46] JJ Dong, B Schmittmann, and RKP Zia. Towards a model for protein production rates. *Journal of Statistical Physics*, 128(1):21–34, 2007.

[47] JJ Dong, RKP Zia, and B Schmittmann. Understanding the edge effect in TASEP with mean-field theoretic approaches. *Journal of Physics A-Mathematical and Theoretical*, 42(1):015002, 2009.

[48] JS Dong, R Lai, K Nielsen, CA Fekete, HF Qiu, and AG Hinnebusch. The essential ATP-binding cassette protein RLI1 functions in translation by promoting preinitiation complex assembly. *Journal of Biological Chemistry*, 279(40):42157–42168, 2004.

[49] DA Doyle, JM Cabral, RA Pfuetzner, AL Kuo, JM Gulbis, SL Cohen, BT Chait, and R MacKinnon. The structure of the potassium channel: molecular basis of K conduction and selectivity. *Science*, 280(5360):69–77, 1998.

[50] DA Drew. A mathematical model for prokaryotic protein synthesis. *Bulletin of Mathematical Biology*, 63(2):329–351, 2001.

[51] RD Driver. *Ordinary and Delay Differential Equations*. *Springer-Verlag*, 1977.

[52] MB Elowitz, AJ Levine, ED Siggia, and PS Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[53] M Fannes, B Nachtergaele, and RF Werner. Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics*, 144(3):443–490, 1992.

[54] G Felsenfeld and GL Cantoni. Use of thermal denaturation studies to investigate the base sequence of *yeast serine* sRNA. *Proceedings of the National Academy of Sciences*, 51(5):818–826, 1964.

[55] L Flintoft. Gene regulation: in charge of translation. *Nature Reviews Genetics*, 14(5):304, 2013.

[56] ME Foulaadvand, AB Kolomeisky, and H Teymouri. Asymmetric exclusion processes with disorder: Effect of correlations. *Physical Review E*, 78(6):061116, 2008.

[57] S Franckenberg, T Becker, and R Beckmann. Structural view on recycling of archaeal and eukaryotic ribosomes after canonical termination and ribosome rescue. *Current Opinion in Structural Biology*, 22(6):786–796, 2012.

[58] CS Fraser, JY Lee, GL Mayeur, M Bushell, JA Doudna, and JWB Hershey. The j-subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40S ribosomal subunits in vitro. *Journal of Biological Chemistry*, 279(10):8946–8956, 2004.

[59] SL French, TJ Santangelo, AL Beyer, and JN Reeve. Transcription and translation are coupled in *archaea*. *Molecular Biology and Evolution*, 24(4):893–895, 2007.

[60] DR Gallie. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes & Development*, 5(11):2108–2116, 1991.

[61] DR Gallie. A tale of two termini: a functional interaction between the termini of an mRNA is a prerequisite for efficient translation initiation. *Gene*, 216(1):1–11, 1998.

[62] A Garai, D Chowdhury, D Chowdhury, and TV Ramakrishnan. Stochastic kinetics of ribosomes: single motor properties and collective behavior. *Physical Review E*, 80(1):011908, 2009.

[63] A Garai, D Chowdhury, and TV Ramakrishnan. Fluctuations in protein synthesis from a single RNA template: stochastic kinetics of ribosomes. *Physical Review E*, 79(1):011916, 2009.

[64] MA Gilchrist and A Wagner. A model of protein translation including codon bias, nonsense errors, and ribosome recycling. *Journal of Theoretical Biology*, 239(4):417–434, 2006.

[65] DT Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403 – 434, 1976.

[66] DT Gillespie. Exact stochastic simulation of coupled chemical-reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[67] P Greulich, L Ciandrini, RJ Allen, and MC Romano. Mixed population of competing totally asymmetric simple exclusion processes with a shared reservoir of particles. *Physical Review E*, 85(1):011142, 2012.

[68] P Greulich and A Schadschneider. Single-bottleneck approximation for driven lattice gases with disorder and open boundary conditions. *Journal of Statistical Mechanics-Theory and Experiment*, 2008(4):P04009, 2008.

[69] JS Griffith. Mathematics of cellular control processes i. negative feedback to one gene. *Journal of Theoretical Biology*, 20(2):202, 1968.

[70] A Griffiths, SR Wessler, RC Lewontin, and SB Carroll. *Introduction to Genetic Analysis*, ninth edition. *W. H. Freeman*, 2007.

[71] S Gu, L Jin, F Zhang, P Sarnow, and A Kay. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nature Structural & Molecular Biology*, 16(2):144–150, 2009.

[72] V Hakim and JP Nadal. Exact results for 2D directed anumals on a strip of finite width. *Journal of Physicas A-Mathematical and General*, 16(7):L213–L218, 1983.

[73] JK Hale, H Buttanri, and H Kocak. *Dynamics and Bifurcations*. *Springer New York*, 2012.

[74] RJ Harris and RB Stinchcombe. Disordered asymmetric simple exclusion process: mean-field treatment. *Physical Review E*, 70(1):016108, 2004.

[75] R Heinrich and TA Rapoport. Mathematical modelling of translation of mRNA in eucaryotes: steady states, time-dependent processes and application to reticulocytes. *Journal of Theoretical Biology*, 86(2):279–313, 1980.

[76] C Hertweck. Biosynthesis and charging of pyrrolysine, the 22nd genetically encoded amino acid. *Angewandte Chemie*, 50(41):9540–9541, 2011.

[77] A Heyd and DA Drew. A mathematical model for elongation of a peptide chain. *Bulletin of Mathematical Biology*, 65(6):1095–1109, 2003.

[78] DJ Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[79] DJ Higham. Modeling and simulating chemical reactions. *SIAM Review*, 50(2):347–368, JUN 2008.

[80] B Hille. *Ion Channels of Excitable Membranes*, thirrd edition. *Sinauer Associates Inc*, 2001.

[81] E Hornstein, H Harel, G Levy, and O Meyuhas. Overexpression of poly(A)-binding protein down-regulates the translation or the abundance of its own mRNA. *FEBS Letters*, 457(2):209–213, 1999.

[82] C Jain. Degradation of mRNA in *Escherichia coli*. *IUBMB Life*, 54(6):315–321, 2002.

[83] M Jansen and P Pfaffelhuber. Stochastic gene expression with delay. *Journal of Theoretical Biology*, 364:355–363, 2015.

[84] J Johansson, P Mandin, A Renzoni, C Chiaruttini, M Springer, and P Cossart. An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell*, 110(5):551–561, 2002.

[85] A Kahvejian, YV Svitkin, R Sukarieh, MN M'Boutchou, and N Sonenberg. Mammalian poly(A)-binding protein is a eukaryotic translation initiation factor, which acts via multiple mechanisms. *Genes & Development*, 19(1):104–113, 2005.

[86] A Klumper, A Schadschneider, and J Zittartz. Equivalence and solution of anisotropic spin-1 models and generalized t-J fermion models in one dimension. *Journal of Physicas A-Mathematical and General*, 24(16):L955–L959, 1991.

[87] AB Kolomeisky. Asymmetric simple exclusion model with local inhomogeneity. *Journal of Physics A-Mathematical and General*, 31(4):1153–1164, 1998.

[88] G Korkmaz, M Holm, T Wiens, and S Sanyal. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *Journal of Biological Chemistry*, 289(44):30334–30342, 2014.

[89] M Kozak. Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2):187–208, 1999.

[90] J Krug. Phase separation in disordered exclusion models. *Brazilian Journal of Physics*, 30(1):97–104, 2000.

[91] U Kuhn and E Wahle. Structure and function of poly(A) binding proteins. *Biochimical et Biophysica Acta - Gene Structure and Expression*, 1678(2-3):67–84, 2004.

[92] S Laalami, L Zig, and H Putzer. Initiation of mRNA decay in bacteria. *Cellular and Molecular Life Sciences*, 71(10):1799–1828, 2014.

[93] G Lakatos and T Chou. Totally asymmetric exclusion processes with particles of arbitrary size. *Journal of Physics A-Mathematical and General*, 36(8):2027–2041, 2003.

[94] B Lewin. *Genes: VII. Oxford University Press, Incorporated*, 2000.

[95] K Li, R Kisilevs, MT Wasan, and G Hammond. Compter simulation of *in vivo* protein synthesis. *Biochimical et Biophysica Acta*, 272(3):451, 1972.

[96] ST Liang, YC Xu, P Dennis, and H Bremer. mRNA composition and control of bacterial gene expression. *Journal of Bacteriology*, 182(11):3037–3044, 2000.

[97] H Lodish. *Cell Biology. W. H. Freeman*, 2008.

[98] CT MacDonald and JH Gibbs. Concerning the kinetics of polypeptide synthesis on polyribosomes. *Biopolymers*, 7(5):707–725, 1969.

[99] CT MacDonald, JH Gibbs, and AC Pipkin. Kinetics of biopolymerization on nucleic acid templates. *Biopolymers*, 6(1):1–25, 1968.

[100] R MacKinnon, SL Cohen, AL Kuo, A Lee, and BT Chait. Structural conservation in prokaryotic and eukaryotic potassium channels. *Science*, 280(5360):106–109, 1998.

[101] N Majdalani, C Cunning, D Sledjeski, T Elliott, and S Gottesman. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21):12462–12467, 1998.

[102] N Malys and J McCarthy. Translation initiation: variations in the mechanism can be anticipated. *Cellular and Molecular Life Sciences*, 68(6):991–1003, 2011.

[103] M Margaliot and T Tuller. Ribosome flow model with positive feedback. *Journal of the Royal Society Interface*, 10(85), 2013.

[104] E Marshall, I Stansfield, and MC Romano. Ribosome recycling induces optimal translation rate at low ribosomal availability. *Journal of the Royal Society Interface*, 11(98), 2014.

[105] W Martin and EV Koonin. Introns and the origin of nucleus-cytosol compartmentalization. *Nature*, 440(7080):41–45, 2006.

[106] A Moghal, K Mohler, and M Ibba. Mistranslation of the genetic code. *FEBS Letters*, 588(23):4305–4310, 2014.

[107] N Morozova, A Zinovyev, N Nonne, L Pritchard, AN Gorban, and A Harel-Bellan. Kinetic signatures of microRNA modes of action. *RNA-A Publication of the RNA Society*, 18(9):1635–1655, 2012.

[108] A Nagar, A Valleriani, and R Lipowsky. Translation by ribosomes with mRNA degradation: exclusion processes on aging tracks. *Journal of Statistical Physics*, 145(5):1385–1404, 2011.

[109] T Nakamoto. Evolution and the universality of the mechanism of initiation of protein synthesis. *Gene*, 432(1):1–6, 2009.

[110] OPD Neto, N Standart, and CM Desa. Autoregulation of poly(A)-binding protein synthesis *in vitro*. *Nucleic Acids Research*, 23(12):2198–2205, 1995.

[111] E Nudler and AS Mironov. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*, 29(1):11–17, 2004.

[112] DJ Obbard, KHJ Gordon, AH Buck, and FM Jiggins. The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society Biological Sciences*, 364(1513):99–115, 2009.

[113] L Pauling, RB Corey, and HR Branson. The structure of proteins - 2 hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, 37(4):205–211, 1951.

[114] RS Pillai, SN Bhattacharyya, CG Artus, T Zoller, N Cougot, E Basyuk, E Bertrand, and W Filipowicz. Inhibition of translational initiation by *Let-7* microRNA in human cells. *Science*, 309(5740):1573–1576, 2005.

[115] AV Pisarev, MA Skabkin, VP Pisareva, OV Skabkina, AM Rakotondrafara, MW Hentze, CU. T Hellen, and TV Pestova. The role of *ABCE*1 in eukaryotic posttermination ribosomal recycling. *Molecular Cell*, 37(2):196–210, 2010.

[116] C Portier, L Dondon, and M Grunbergmanago. Translational autocontrol of the *Escherichia coli* ribosomal protein S15. *Journal of Molecular Biology*, 211(2):407–414, 1990.

[117] E Redon, P Loubiere, and M Cocaign-Bousquet. Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *Journal of Bilogical Chemistry*, 280(43):36380–36385, 2005.

[118] T Reichenbach, T Franosch, and E Frey. Exclusion processes with internal states. *Physical Review Letters*, 97(5), 2006.

[119] AS Ribeiro. Stochastic and delayed stochastic models of gene expression and regulation. *Mathematical Biosciences*, 223(1):1–11, 2010.

[120] MV Rodnina, M Beringer, and W Wintermeyer. How ribosomes make peptide bonds. *Trends in Biochemical Sciences*, 32(1):20–26, 2007.

[121] MC Romano, M Thiel, I Stansfield, and C Grebogi. Queueing Phase Transition: Theory of Translation. *Physical Review Letters*, 102(19), 2009.

[122] AK Sharma and D Chowdhury. Stochastic theory of protein synthesis and polysome: Ribosome profile on a single mRNA transcript. *Journal of Theoretical Biology*, 289:36–46, 2011.

[123] LB Shaw, JP Sethna, and KH Lee. Mean-field approaches to the totally asymmetric exclusion process with quenched disorder and large particles. *Physical Review E*, 70(2):021901, 2004.

[124] LB Shaw, RKP Zia, and KH Lee. Totally asymmetric exclusion process with extended objects: a model for protein synthesis. *Physical Review E*, 68(2):021910, 2003.

[125] CJ. Shoemaker and R Green. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in *yeast*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(51):1392–1398, 2011.

[126] M Siwiak and P Zielenkiewicz. A Comprehensive, quantitative, and genome-wide model of translation. *PLoS Computational Biology*, 6(7):e1000865, 2010.

[127] N Skjondal-Bar and DR Morris. Dynamic model of the process of protein synthesis in eukaryotic cells. *Bulletin of Mathematical Biology*, 69(4):1451, 2007.

[128] N Sonenberg and AG Hinnebusch. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell*, 136(4):731–745, 2009.

[129] C Steglich, D Lindell, M Futschik, T Rector, R Steen, and SW Chisholm. Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biology*, 11(5):R54, 2010.

[130] IS Tolokh, S Goldman, and CG Gray. Unified modeling of conductance kinetics for low- and high-conductance potassium ion channels. *Physical Review Letters*, 74(1), 2006.

[131] H Tourriere, K Chebli, and J Tazi. mRNA degradation machines in eukaryotic cells. *Biochimie*, 84(8):821–837, 2002.

[132] G Tripathy and M Barma. Driven lattice gases with quenched disorder: exact results and different macroscopic regimes. *Physical Review E*, 58(2):1911–1926, 1998.

[133] T Tuller and E Kupiec, Mand Ruppin. Determinants of protein abundance and translation efficiency in *S-cerevisiae*. *PLoS Computational Biology*, 3(12):2510–2519, 2007.

[134] JJ Tyson, KC Chen, and B Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003.

[135] A Valleriani, Z Ignatova, A Nagar, and R Lipowsky. Turnover of messenger RNA: polysome statistics beyond the steady state. *Europhysics Letters*, 89(5), 2010.

[136] G Vassart, JE Dumont, and Cantrain.FR. Translational control of protein synthesis - a simulation study. *Biochimical et Biophysica Acta*, 247(3):471, 1971.

[137] D Vautherin and DM Brink. Hartree-Fock calculations with Skyrme's Interaction. I. Spherical nuclei. *Phys. Rev. C*, 5:626–647, 1972.

[138] T Von der Haar. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC Systems Biology*, 2, 2008.

[139] EGH Wagner, S Altuvia, and P Romby. Antisense RNAs in bacteria and their genetic elements. *Homology Effects*, 46:361–398, 2002.

[140] B Wang, TM Love, ME Call, JG Doench, and CD Novina. Recapitulation of short RNA-directed translational gene silencing *in vitro*. *Molecular Cell*, 22(4):553–560, 2006.

[141] W Wang and DH Bechhofer. Properties of a *Bacillus subtilis* polynucleotide phosphorylase deletion strain. *Journal of Bacteriology*, 178(8):2375–2382, 1996.

[142] DJ Wilkinson. Stochastic modelling for systems biology. *Taylor & Francis*, 2006.

[143] E WORK. A new naturally occurring amino-acid. *Nature*, 165(4185):74–75, 1950.

[144] J Wu and J Bag. Negative control of the poly(A)-binding protein mRNA translation is mediated by the adenine-rich region of its 5' untranslated region. *Journal of Biological Chemistry*, 273(51):34535–34542, 1998.

[145] LG Wu, JH Fan, and JG Belasco. MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11):4034–4039, 2006.

[146] A Yarunin, VG Panse, E Petfalski, C Dez, D Tollervey, and EC Hurt. Functional

link between ribosome formation and biogenesis of iron-sulfur proteins. *EMBO Journal*, 24(3):580–588, 2005.

[147] T You, GM Coghill, and AJP Brown. A quantitative model for mRNA translation in *Saccharomyces cerevisiae*. *Yeast*, 27(10):785–800, 2010.

[148] R Young and H Bremer. Polypeptide-chain-elongation rate in *Escherichia coli* as a function of growth rate. *Biochemical Journal*, 160(2):185–194, 1976.

[149] J Yu and JE Russell. Structural and functional analysis of an mRNP complex that mediates the high stability of human beta-globin mRNA. *Mollecular and Cellular Biology*, 21(17):5879–5888, 2001.

[150] R Zhu, AS Ribeiro, D Salahub, and SA Kauffman. Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. *Journal of Theoretical Biology*, 246(4):725–745, 2007.

[151] R Zhu and D Salahub. Delay stochastic simulation of single-gene expression reveals a detailed relationship between protein noise and mean abundance. *FEBS Letters*, 582(19):2905–2910, 2008.

[152] RKP Zia, JJ Dong, and B Schmittmann. Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *Journal of Statistical Physics*, 144(2):405–428, 2011.

[153] A Zinovyev, N Morozova, N Nonne, E Barillot, A Harel-Bellan, and AN Gorban. Dynamical modeling of microRNA action on the protein translation process. *BMC Systems Biology*, 4, 2010.

[154] H Zouridis and V Hatzimanikatis. A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. *Biophiscial Journal*, 92(3):717–730, 2007.